# Aversive Reinforcement Learning

**Ben Seymour**

**Wellcome Trust Centre for Neuroimaging, UCL**
**12 Queen Square,**
**London WC1N 3BG**

**Supervisors:**
**Ray Dolan**
**Richard Frackowiak**
**Karl Friston**

**Submitted for the consideration of a PhD in Neurological Science.**

## Declaration.

**I, Ben Seymour, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.**

## Abstract.

**We hypothesise that human aversive learning can be described algorithmically by Reinforcement Learning models. Our first experiment uses a second-order conditioning design to study sequential outcome prediction. We show that aversive prediction errors are expressed robustly in the ventral striatum, supporting the validity of temporal difference algorithms (as in reward learning), and suggesting a putative critical area for appetitive-aversive interactions. With this in mind, the second experiment explores the nature of pain relief, which as expounded in theories of motivational opponency, is rewarding. In a Pavlovian conditioning task with phasic relief of tonic noxious thermal stimulation, we show that both appetitive and aversive prediction errors are co-expressed in anatomically dissociable regions (in a mirror opponent pattern) and that striatal activity appears to reflect integrated appetitive-aversive processing. Next we designed a Pavlovian task in which cues predicted either financial gains, losses, or both, thereby forcing integration of both motivational streams. This showed anatomical dissociation of aversive and appetitive predictions along a posterior-anterior gradient within the striatum, respectively.**

**Lastly, we studied aversive instrumental control (avoidance). We designed a simultaneous pain avoidance and financial reward learning task, in which subjects had to learn independently learn about each, and trade off aversive**

**and appetitive predictions. We show that predictions for both converge on the medial head of caudate nucleus, suggesting that this is a critical site for appetitive-aversive integration in instrumental decision making. We also study also tested whether serotonin (5HT) modulates either phasic or tonic opponency using acute tryptophan depletion. Both behavioural and imaging data confirm the latter, in which it appears to mediate an average reward term, providing an aspiration level against which the benefits of exploration are judged.**

**In summary, our data provide a basic computational and neuroanatomical framework for human aversive learning. We demonstrate the algorithmic and implementational validity of reinforcement learning models for both aversive prediction and control, illustrate the nature and neuroanatomy of appetitive-aversive integration, and discover the critical (and somewhat unexpected) central role for the striatum.**

# Ackowledgements.

**I would like to thank my supervisors Karl Friston, Richard Frackowiak and Ray Dolan for their help and support. I would also especially like to thank John O’Doherty, Nathaniel Daw and Peter Dayan for their substantial scientific guidance and mentorship throughout my doctoral research. I also thank a number of valued collaborators, including Chris Frith, Martin Koltzenburg, Anthony Jones, Katya Wiech, Tania Singer, Jon Roiser, Ivo Vlaev and Nick Chater, and many other fellow fellows at the Functional Imaging Lab and broader UCL neuroscience community.**

**I would also like to thank the Wellcome Trust for their financial support, and the Gatsby Charitable Foundation for supporting the ‘Pain and Learning’ workshop in 2006, organised with by myself, Martin Koltzenburg and Peter Dayan.**

# Contents.

## Publications arising directly from this submission.

1. Ben Seymour, John O'Doherty, Peter Dayan, et al Temporal difference models describe higher-order learning in humans. *Nature*. 2004 Jun 10;429(6992):664-7. (Chapter 3)

2. Ben Seymour, John O'Doherty, Martin Koltzenburg, et al. Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nature Neuroscience*. 2005 Sep;8(9):1234-40. (Chapter 4).

3. Ben Seymour, Tania Singer, Ray Dolan. The neurobiology of punishment. *Nature Reviews Neuroscience* 2007 8; 300-11. (Chapter 1 and 7).

4. Ben Seymour, Nathaniel Daw, Peter Dayan, Tania Singer, Ray Dolan. Differential responses to gains and losses in human striatum. *Journal of Neuroscience*. 2007 May 2;27(18):4826-31.

5. Peter Dayan, Ben Seymour. Values and actions in aversion. In N*euroeconomics: decision making in the brain* Eds. Glimcher, Fehr, Camerer and Poldrack. Elsevier 2008. (Chapter 1 and 7).

6. Ben Seymour and Ray Dolan. Emotion, decision making and the amygdala. *Neuron* June 12th 2008. (Chapter 1 and 7).

7. Ben Seymour and Sam McClure. Anchors, scales and relative coding of value in the brain. *Current Opinion in Neurobiology*. 2008 Apr;18(2):173-8. (Chapter 7).

8. Ivo Vlaev and Ben Seymour, Ray Dolan, Nick Chater. The price of pain and the value of suffering. *Psychological Science,* 2009. (Chapter 7).

9. Ben Seymour, Wako Yoshida, Ray Dolan. Altruistic Learning. *Frontiers in Neuroscience,* 23:3 2009. (Chapter 7).

10. Ben Seymour, Nathaniel Daw, Peter Dayan, Jon Roiser, Karl Friston, Ray Dolan. Serotonin mediates tonic appetitive-aversive interactions in human striatum. Submitted. (chapter 6).

# Figures.

## 1.1 Summary:

**Aversive events, such as pain, are characterised (and arguably defined) by the property by which they induce behaviour that attempts to reduce or terminate their current and future occurrence. This thesis studies how this is achieved in humans. We take a Marrian approach, and first formalise the problem as one of optimal control. We propose that aversive learning can be understood theoretically as a Reinforcement Learning problem. The majority of our work is built on this framework and tests predictions that come from the behavioural and implementational hypotheses that it derives. These hypotheses manifest algorithmically as temporal difference learning and Q learning models, based on their simplicity, biologically plausibility, and emerging parallel evidence from studies of appetitive learning.**

**The first experiment tests a basic prediction of Reinforcement Learning algorithms, that prediction errors are expressed somewhere in the brain. We use Pavlovian aversive conditioning using visual cues and painful shocks, and study brain activity using parametric fMRI. In particular, we use a second-order conditioning design to look for evidence of higher order prediction errors, which are a specific prediction of temporal difference learning, since they reveal the mechanism by which predictive values are 'bootstrapped' between sequential cues. We show that prediction errors are expressed, most robustly in the ventral striatum, providing good support for the validity of temporal difference algorithms. What is surprising is the role it suggests for the striatum, given its reputation as a reward-specific area – previous studies have shown expression of appetitive prediction errors in precisely the same region. This relationship between reward and aversion becomes a dominant theme in subsequent experiments.**

**The second experiment explores the nature of pain relief. Pain relief is inherently rewarding, and illustrates the excitatory-inhibitory opponent relationship between rewards and punishments. Relief can be achieved either by omission of an otherwise expected phasic aversive event, or by**

**termination of a tonic aversive event. This can be fit easily within a Reinforcement Learning framework by postulating the existence of distinct appetitive and aversive learning systems. We designed a Pavlovian conditioning task with tonic thermal stimulation of capsaicin-sensitised skin (a good physiological model of injury), in which visual cues predicted either phasic exacerbation of tonic pain, or relief of pain (induced by transient cooling of the skin). We show that both appetitive and aversive prediction errors are co-expressed in anatomically dissociable regions, and in a mirror opponent pattern. Whereas the amygdala adopts a reward specific role, and the lateral orbitofrontal cortex an aversive-specific role, striatal activity appears to reflect integrated appetitive-aversive activity, showing an interaction between valence and cue type (excitatory or inhibitory): that is, it shows positive prediction error activity for excitatory cues, regardless of valence.**

**Whereas the preceding experiment provides good evidence for the implementation of opponent temporal difference learning, it raises important questions about exactly what is being processed in the striatum, since it remains to be shown that it can even distinguish rewards from punishments. The next experiment explores this in more detail. We designed a probabilistic Pavlovian conditioning task in which visual cues predicted either financial gains, losses, or both. This ought to force the striatum to integrate both motivational streams, since the prediction error need be constructed by both. This imaging data showed anatomical dissociation of aversive and appetitive predictions, with aversive prediction errors being expressed more posterior and dorsal to appetitive prediction errors, consistent with recent stimulation studies illustrating a valence gradient of motivational behaviour in rodent striatum. This experiment also illustrates the computational and anatomical similarity between secondary punishments – financial loss, and primary punishments such as pain.**

**The above studies provide a solid computational and anatomical account of aversive Pavlovian learning. However, ignoring for a moment the influence of conditioned responses (such as withdrawal), Pavlovian learning permits**

**only *prediction* and not *control*. In both animal learning (operant / instrumental learning) and computational Reinforcement learning theory (eg. Q learning), the latter can be achieved by learning the outcomes of specific actions, which allow them to manipulate their environment to minimise (escape future threats). In instrumental avoidance learning paradigms, the avoidance state is known to act as a reward (a conditioned inhibitor) that can reinforce actions. In the next experiment, we studied first whether reinforcement learning (Q learning) can capture instrumental avoidance, and how this is represented in the brain, in relation to simple reward based instrumental learning. To do this, we designed a simultaneous pain avoidance and financial reward learning task, which given the two outcomes were independently contingent on actions, forces the individual to learn about each separately, and trade off aversive (pain) and appetitive (money) outcome predictions. This showed that predictions for both converge on the medial head of caudate nucleus, suggesting that this is a critical site for appetitive-aversive integration in instrumental decision making.**

**The above study also tested one further hypothesis: whether 5HT (serotonin) mediates either phasic opponency between punishments and rewards, or tonic opponency between phasic and tonic outcomes. Using acute dietary tryptophan depletion to manipulate central serotonin: we show that learning from either rewards or punishments, and the trade-off between the two, are not substantially influenced by tryptophan status. However, we show that the representation of average reward, which acts as a tonic signal against which phasic rewards are compared (tonic opponency), is significantly decreased in the depleted group. This manifests in a subjects' tendency to maintain responding for previous actions ('choice stickiness'). In the brain, we show that phasic reward and punishment related activity converge on the medial head of caudate nucleus in the basal ganglia, but that activity here associated with choice stickiness positively correlates with serotonin, indicating that the modification of value according to average reward occurs outside of the caudate. These data suggest a specific computational account in which serotonin controls an average**

reward signal against which any choice’s outcomes are weighed, and illustrates the integrative but partial role of the caudate nucleus in computing values associated with choice.

In summary, our data provide a basic computational and neuroanatomical framework for human aversive learning. We demonstrate the algorithmic and implementational validity of reinforcement learning models for aversive prediction and control, illustrate the nature and neuroanatomy of appetitive-aversive integration, and discover the critical (and somewhat unexpected) central role for the striatum. We discuss the broader implications of these results for decision neuroscience, behavioural economics and social neuroscience.

## 1.2 Aversive animal learning theory.

### 1.2.1 Innate and Pavlovian value.

Our current understanding of aversive motivation owes much to many decades of invaluable work by experimental psychologists, and to the many thousands of brave rodents and other animals who have assisted them in their pursuits. It is the basic tenets of animal learning theory that provide the framework for the current work, which aims to explore the neurobiological basis of aversive motivation and decision-making in humans.

There are two fundamental components to motivation. The first is action. Actions allow us either to increase the probability of a rewarding outcome (appetitive motivation), or to reduce the probability of a punishing outcome (aversive motivation). From a motivational perspective, a reward can be defined as an event that an animal will expend energy to bring about, whereas an aversive event (punishment) is something an animal will expend energy to reduce or avoid.

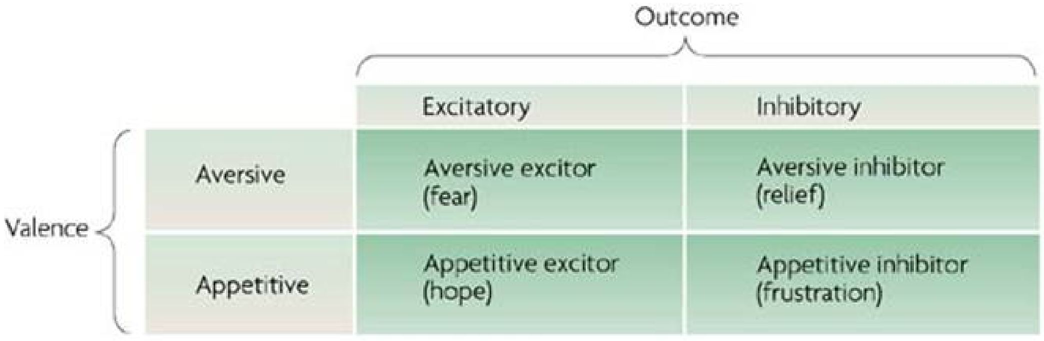
The second component is learning. Actions that result in a higher than expected reward, or lower than expected punishment are *reinforced* – they are more likely to be reproduced in a similar situation again. Actions that result in lower than expected reward, or greater than expected punishment have the opposite effect, being less likely to be produced again. It is this comparison between expected and actual outcomes that seeds one of the fundamental concepts of both animal learning theoretic and computational approaches to learning: that errors in prediction should be a useful quantity in guiding future action (Rescorla and & Wagner, 1972).

The two basic types of learning paradigm inherited from experimental psychology - Pavlovian and instrumental conditioning – reflect an important distinction in theoretical approaches to motivation. Pavlovian conditioning establishes statistically predictive pairings between environmental cues (‘conditioned stimuli’) and salient outcomes (‘unconditioned stimuli’ such as shocks or food), regardless of any action the individual can elicit. That is, although the cue will come to elicit a conditioned response, this response does not change the probability of the outcome. The fact that conditioned responses (such as approach and withdrawal) sometimes do change the probability of an outcome both justifies their evolution, and confounds experiments (when it comes to interpreting data from a range of paradigms such as autoshaping, conditioned place aversion, escape learning for example). Aside from this, however, Pavlovian conditioning is primarily concerned with *prediction*, and the magnitude of the conditioned response reflects (ie. is some function of) the magnitude of the predicted outcome.

Instrumental learning establishes the statistical association between an action and an outcome. That is, elicitation of a specific outcome *does* change the probability of an action, in contrast to Pavlovian conditioning. In this way, an animal can accrue rewards and avoid punishments by learning to perform certain actions rather than others, when in a particular situation (defined by the ‘discriminative stimulus’). In this way, instrumental learning permits *control* of the environment.

At the heart of attempts to formalise motivation is the concept of value. For instance, the value of an aversive event can be considered in terms of an ordinal scale of preference, in which, given a choice, less aversive outcomes will be selected over higher aversively-valued ones. This concept of value is useful, as it outlines a unitary currency against which events of different modalities can be judged (Montague and Berns, 2002). Through learning, value incorporates otherwise neutral states or cues that *predict* ‘primary’ rewards or aversive events to some degree. To reiterate, this ‘state–outcome’ associability is embodied within Pavlovian learning, in which a reliable predictive pairing of the conditioned stimulus with an unconditioned stimulus. For example, being bitten by a particular dog is likely to induce increased heart rate, sweating and fleeing when that dog is encountered subsequently. The aversive value of the dog reflects the severity of its bite.

The ability to predict aversive events (and rewards) has self-evident motivational benefits, but also yields a new set of possible events, namely those associated with omission of an expected outcome. Accordingly, the omission of an expected aversive event can be rewarding (an aversive inhibitor), and the omission of reward can be aversive (an appetitive inhibitor). This relationship underpins a basic architecture of motivational systems in which reward and aversive mechanisms oppose each other (Dickinson and Dearing, 1979;Konorski, 1967). This ‘Konorskian’ model consists of underlying, mutually inhibitory appetitive and aversive systems whose operation gives rise to four basic categories of motivation – prediction of reward (hope), prediction of aversive events (fear), omission of reward (frustration) and omission of aversive events (relief) (**figure 1.1**; see also (Gray, 1991)).



| | | Outcome | |
|---|---|---|---|
| | | Excitatory | Inhibitory |
| Valence | Aversive | Aversive excitor<br>(fear) | Aversive inhibitor<br>(relief) |
| | Appetitive | Appetitive excitor<br>(hope) | Appetitive inhibitor<br>(frustration) |

Nature Reviews | Neuroscience

**Figure 1.1** Motivational stimuli can be excitatory or inhibitory, depending on whether they predict the occurrence or the absence, respectively, of an affective outcome or of another predictor. They can also be classified by valence, as stimuli that are associated with either appetitive or aversive outcomes or predictors. When combined, these two classifications illustrate the four basic motivational states of fear, relief, hope and frustration (figure from (Seymour et al., 2007b)).

The reciprocity between appetitive and aversive motivational systems was demonstrated in a series of elegant experiments, depicted in **figure 1.2,** below**.** In the paradigm termed ‘blocking’ **(a),** there is a failure of a novel cue to acquire an aversive conditioned response to an outcome that is already well predicted by an existing cue (unless it precedes it). This, at the very least phenomenologically, appears to be because there is no ‘aversiveness’ left to predict (Kamin, 1968;Rescorla RA, 1971). In a modification of blocking, termed transreinforcer blocking **(b),** a cue that already predicts an aversive outcome can block the acquisition of a conditioned response to a novel cue that is paired with an aversive outcome in a different modality. For example, a cue that has been pre-trained with a painful foot-shock, presented in compound with a novel cue and paired with a loud aversive noise, blocks conditioning to the novel cue (Kamin et al., 1963). Even though noise and pain differ in their sensory properties, they seem to access a common aversive system, indicating that punishments of any modality might be treated in a similar way.

But Dickenson and Dearing provided a final, ingenious twist to transreinforcer blocking **(c):** they wondered whether it could be accomplished by using a conditioned inhibitor. In conditioned inhibition, a cue that predicts that an otherwise expected reward will be omitted (causing frustration) acquires aversive properties: for instance, it will suppress instrumental appetitive responding (conditioned suppression(Bull and Overmier, 1968)), and be slow to acquire future conditioned responses to a reward (retardation (Rescorla, 1969)). Dickinson and Dearing showed that a cue that signals the omission of food pellets will block a primary aversive punisher (Dickinson and Dearing MF, 1979). Rats were pre-trained with a cue that was unpaired with food (and therefore acted as an appetitive conditioned inhibitor), and this cue was subsequently presented in compound with a novel cue, and followed by a painful foot-shock **(d).** Testing the value of this new cue (by conditioned suppression), in comparison to controls, showed that the conditioned inhibitor for food successfully blocked prediction of the foot-shock. This provides critical evidence to support the existence of a common underlying Pavlovian aversive representation.

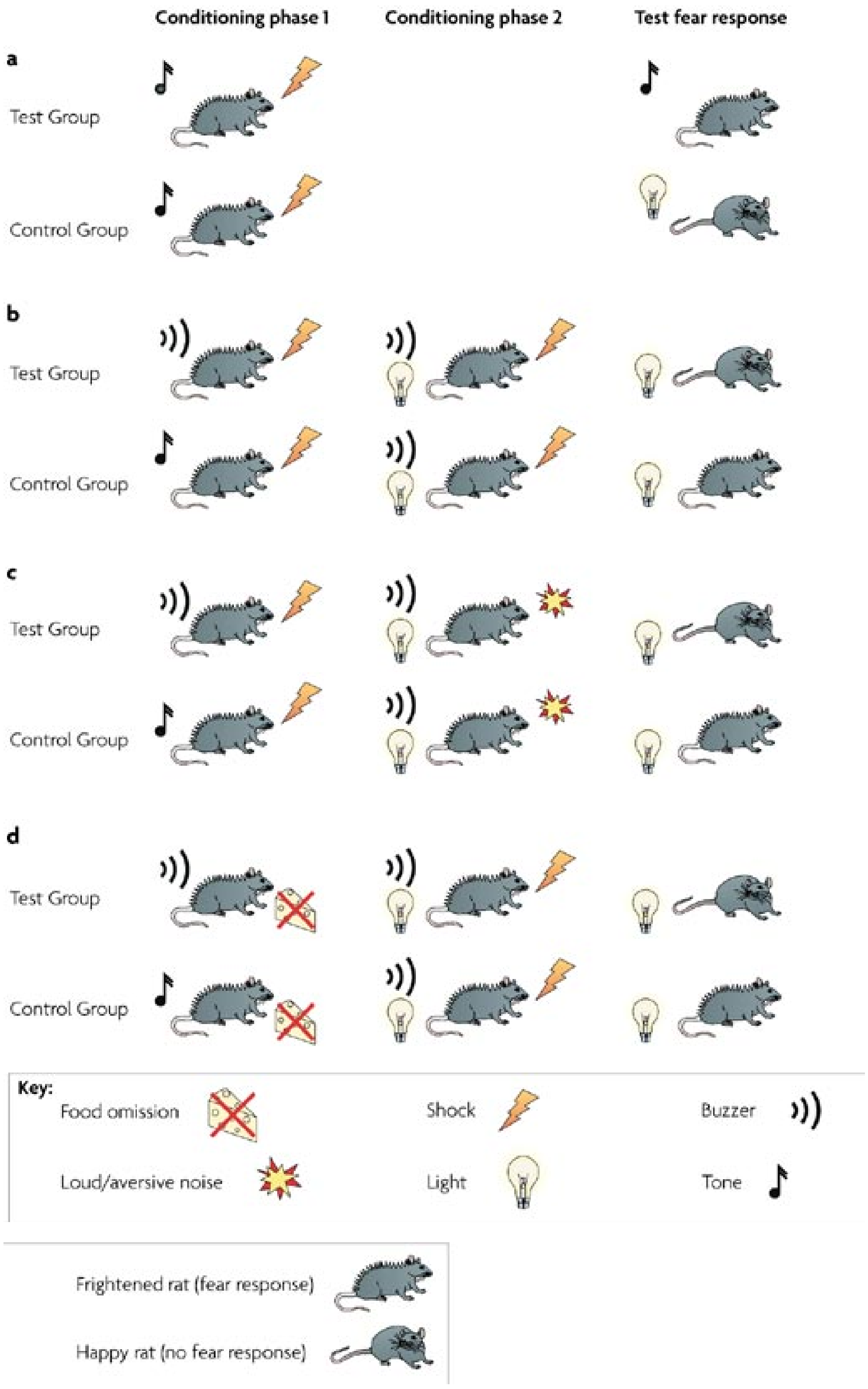


**Figure 1.2** Learning paradigms that have helped reveals the underlying structure of appetitive and aversive motivational systems. Panels are referred to in the text above and below. Figure from (Seymour et al., 2007b).

There are two important caveats to this. The first caveat is that predictive representations, inferred by the nature of responses they induce and their properties in experiments like that above, fall into two categories. First are general motivational, stimulus non-specific representations: it is this representation that is captured by the blocking experiments above, and produces general conditioned responses such as approach and withdrawal. This representation ignores the identity of the outcome being predicted other than whether it is aversive or appetitive. The second category is stimulus specific representations, which are peculiar and appropriate to the precise nature of outcome. Thus left leg withdrawal is an appropriate response to a cue that predicts painful shock to the left foot, but not to an air-puff to the eye. The basic architecture of these distinct representations is shown in figure 3.

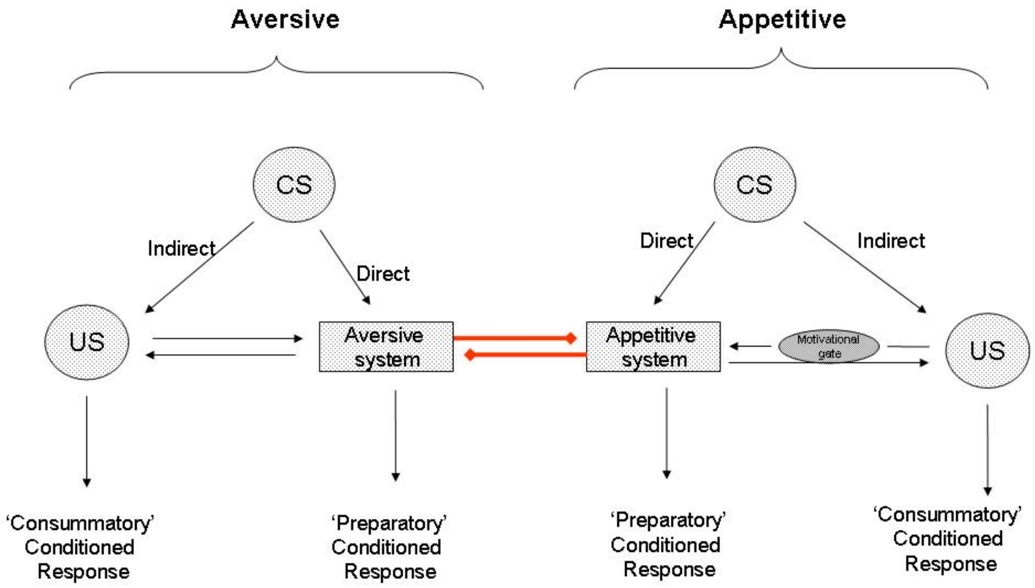

Aversive
Appetitive
CS
Indirect
Direct
US
Aversive system
'Consummatory' Conditioned Response
'Preparatory' Conditioned Response
CS
Direct
Indirect
Appetitive system
Motivational gate
US
'Preparatory' Conditioned Response
'Consummatory' Conditioned Response

Figure 1.3. Konorskian model of Pavlovian appetitive conditioning, showing direct and indirect pathways mediating representation of conditioned stimuli (CS) and unconditioned stimuli (US). Redrawn and adapted from Dickinson and Balleine (2002). See also (Seymour, 2006).

The second caveat is that appetitive-aversive opponency can arise in a related but slightly different circumstance. Konorski's opponency deals with excitators and inhibitors or phasic rewards and punishments. However, elsewhere Solomon and Corbit studied states associated with the offset of tonically presented rewards and punishments. For instance, if my supervisor incentivises my presence in the laboratory with a machine that delivers Maltesers at a rate of 1 per 15mins, then reducing the rate, or stopping it completely, becomes a punishment.

Solomon and Corbit posited the distinction between what they termed A states – the primary excitatory tonic stimulus, and B states – those elicited by the termination of those states (figure 1.4, below). Furthermore, they argued that B 'states', for example the offset relief from tonic aversive stimulation, could independently motivate behviour. A critical feature of Solomon and Corbits thesis related to the temporal behaviour of A and B states, and they suggested that the latter were more resistant to habituation than the former and hence could dominate behaviour. An example they suggested was that skydivers would continue skydiving motivated by the pleasure in relief when their feet were safely on the ground, having habituated to the aversion associated with plummeting towards it at speed. More recently, the same argument has formed a major class of theories of addiction (Koob et al., 1997).

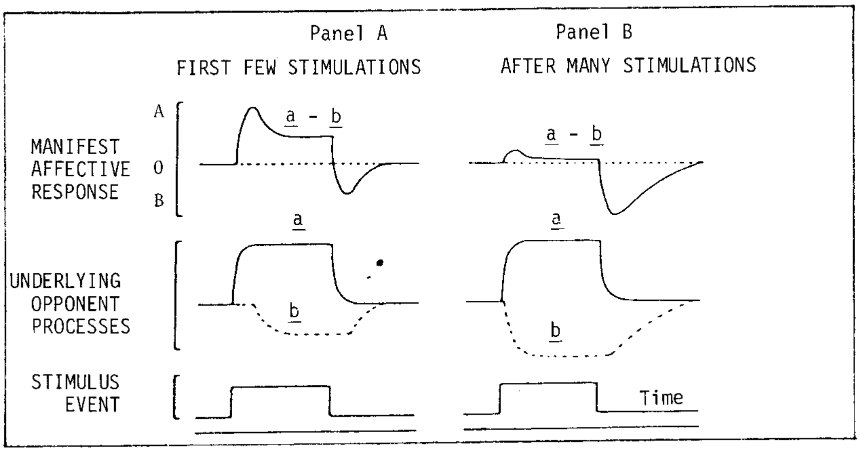


Figure 1.4. Solomon and Corbits Theory of Opponency. This figure, taken directly from their paper, illustrates the operation of excitatory 'a' process and inhibitory 'b' processes. The relative resistance of 'b' processes to extinction allows the offset relief to dominate the motivational value of previously aversive processes, whose excitatory aversive 'a' value have habituated.

Experimental demonstrations of the motivational value of off-set relief are slightly less abundant than theories of its importance, but an elegant example has been shown in *Drosophila*. Tanimoto and colleagues paired an odour cue with the offset of shock, and showed that the odour subsequently attracted the flies when presented alongside a neutral odour (Tanimoto et al., 2004).

In summary, there is reasonable evidence from experimental psychology to support the existence of an underlying general aversive Pavlovian motivational system, which operates alongside and as an opponent to an appetitive system.

### 1.2.2 Action learning.

As hinted above, Pavlovian learning involves slightly more than the acquisition of state values (see figure 3, above), and the responses they elicit can serve an important function. For instance, prediction of aversive events often produces defensive or aggressive responses that clearly evolved to protect the immediate welfare of the animal. Indeed, aggressive responses are often seen towards inanimate aversive cues in animal experiments, these responses can even be elicited by stimuli associated with the omission of food (appetitive inhibitors (Hutchinson et al., 1968)), consistent with an opponent model. Pavlovian aversive actions are often stimulus specific and diverse, indeed far more so than for rewards, involving a wide variety of behaviours ranging from freezing, running, and fighting. They are also often context-dependent: for instance, in a male rat, the prediction of a painful shock may produce freezing in a solitary animal, and aggression in the presence of another male (Ulrich and Azrin, 1962). In addition to the nature of punishment, they are also appropriate for the timing of it, for example eyeblink to an anticipated air-puff is scheduled for the time of the air-puff. This specificity is not exclusively the case, however, as rats will also attack a cue-light that predicts shock, as they will lick and bite it if it predicts reward (stimulus substitution). Thus, the diversity of Pavlovian actions reflects the combination of stimulus specific and non-specific anticipatory actions.

However, Pavlovian actions provide a fundamentally restricted set of options for action, and more flexible control is achieved by instrumental learning, whereby an individual learns to associate a particular action with its outcome (Thorndike, 1911). Consequently, actions that lead to a reward are executed more frequently in future, whereas those that lead to aversive events are executed less often. For example, discovering that pressing a lever results in food delivery will cause an animal to press that lever more often, whereas if such an action is followed by an electric shock, the animal will press the lever less often. A wealth of data has shown that action suppression is proportional to the magnitude, certainty, and imminence of an anticipated punishment (Atnip, 1977;Azrin, 1956;Azrin, 1960;Azrin et al., 1963;Baron, 1965;Camp et al., 1967;Church et al., 1967;Church, 1969a;Solomon et al., 1968;Walters and Grusec, 1977). This effect is in part Pavlovian: cues that were previously paired with punishment suppress instrumental responding in the absence of any instrumental contingency (conditioned suppression (Estes and Skinner, 1941)), but adding such a contingency substantially enhances suppression(Bolles et al., 1980;Church, 1969b).

Instrumental learning allows learning of arbitrary and potentially highly adaptive responses beyond the restrictive set that are available to Pavlovian mechanisms. But instrumental learning is not in itself a unitary process. There are at least two distinct types of instrumental action: habits, and goal-orientated actions. Habits learn the scalar value of actions, by essentially collapsing the value of future outcomes onto a single action-value for each choice available to the animal. Thus, although the (value of the) outcome may be directly used to reinforce, or inhibit, the action, the resulting habit does not encode any specific representation of that outcome, and as such is often regarded as stimulus-response learning (although sometimes stimulus-response learning is taken to involve acquisition of a binary, discrete action, without representation of magnitude).

Habit-based learning may be a highly effective, and computationally simple, way to learn and act following extensive exposure to an environment with predictable outcomes. However, it may be a less effective way to make choices given limited experience, or if the outcomes depend on more complex aspects of the action and the environment. In contrast, goal-orientated actions incorporate an internal representation of the outcome which can be used more directly to guide actions. Experimentally, one of the hallmarks of goal-orientated action is sensitivity to outcome devaluation: if an animal learns to press a lever for food when hungry, and is subsequently fed to satiety, it presses the lever less frequently when exposed to the lever again, indicating that it appropriately represents the reduced value of the action. However, there is good behavioural evidence of a transfer of action control from goal-orientated to habit based systems through time, and on extensive training this sensitivity to outcome devaluation is reduced(Balleine, 2005;Daw et al., 2005;Dickinson and Balleine, 2002).

In addition to simple outcome representations, goal-orientated action selection may accommodate substantial complexity, involving representation of potentially intricate sequences of actions, including those whose outcomes are governed by higher-order structure and rules. Although many animals may possess a surprisingly sophisticated ability to model the structure of their environment to guide goal-orientated behaviour(Blaisdell et al., 2006;Raby et al., 2007), this capacity is clearly remarkably developed in humans. Figure 1.5 illustrates a toy maze based navigation task, and details how different action systems can learn to find reward and avoid punishment.

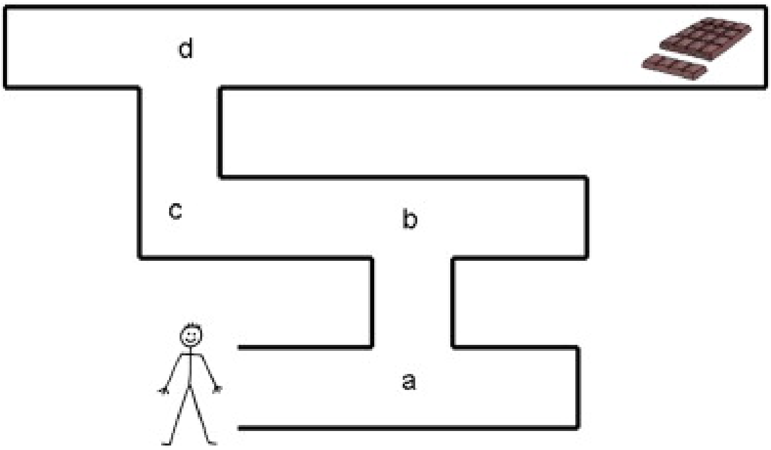


Figure 1.5. **How many action systems?**

Consider the problem of learning to find the food in the maze above. The simplest solution utilises Pavlovian conditioning and exploits innate actions such as approach and withdrawal. During Pavlovian conditioning, positions that are associated with the outcome acquire a positive value that causes the agent to approach them. Thus, following tendency to approach the reward from position **d**, **d** will acquire a positive utility, causing it to be approached **d** from other positions, including **c**. Through sequential conditioning, the individual can potentially navigate relying purely on Pavlovian approach.

Habits involve the learning of action utilities. Trial and error will reveal that turning right at **d** is immediately profitable, and the reward can be used directly to reinforce the action. Learning the preceding actions, such as what to do at position **b** is more difficult, since the outcomes are both delayed and are contingent on subsequent actions (the credit assignment problem (Bellman, 1957)). One possibility is to use either the subsequently available best action utility (as in Q learning (Watkins and Dayan, 1992)), or the subsequent Pavlovian state values (as in Actor-Critic learning (Barto, 1995)), as a surrogate reward indicator. This has the effect of propagating (or 'bootstrapping') action utilities to increasing distances in chains of actions.

Goal directed learning mechanisms overcome the lack of an explicit representation of the structure of the environment or of the utility of a goal in Pavlovian actions and habits, by involving a model of some sort. Indeed, there may be more than one distinct forms of model-based decision system (Yoshida and Ishii, 2006). A natural form is a map of the area within which one's own position and the position of the goal can be specified, in which the structure of the model is governed by the two dimensional physical nature of the environment. Alternatively, propositional models, which have a less constrained prior structure, might specify actions as bringing about transitions between uniquely identified positional states. Such models bear a closer relation to linguistic mechanisms, for instance taking the form of 'from the starting position, go left, left again, then right, and then right again', and in theory have the capacity to incorporate complex sets of state-action rules.

Lastly, control might also be guided by discrete episodic memories of previous reinforcement. Such a controller is based on explicit recall of previous episodes, and has been suggested to guide actions in the very earliest of trials (Lengyel and Dayan, 2007)

### 1.2.3 Avoidance learning

In aversive learning, the most basic instrumental paradigm is avoidance. Typically a subject receives a warning stimulus (such as a tone or light) that precedes delivery of an aversive stimulus, such as prolonged electrification of the floor of one compartment of the experimental apparatus. At first, the subject responds only during the aversive stimulus, for instance escaping the shock by jumping into a neighbouring compartment. Typically, the warning stimulus will be extinguished following this escape response. After several presentations, the escape response is executed more quickly, and eventually, the subject learns to jump when observing the warning stimulus (again with the effect of turning off this stimulus), thus completely avoiding the shock.

Consideration of the problems that must be solved in avoidance hints that such behaviour may not be straightforward. For instance, how are successful avoidance actions reinforced, if by definition they lead to no outcome? (How) does a subject ever realise that the threat is gone, if it is never sampled? Mowrer famously suggested that learning to avoid involves two processes: predicting the threat, and learning to escape from the predictor (Mowrer, 1947). These processes, proposed respectively to be under Pavlovian and instrumental control, comprise two-factor theory, which in one form or another has survived well over the past decades. Although there are many unanswered questions about precisely how the different action systems are orchestrated in different avoidance situations, some key facts are well grounded.

Notably, Pavlovian mechanisms play a critical (and multifarious) role in avoidance, and indeed Pavlovian responses to the warning stimulus alone are often capable of implementing successful avoidance. For example, jumping out of an electrified chamber, blinking in anticipation of an eye-puff, leg flexion to an electric foot plate can all completely remove an aversive stimulus, without any need for an instrumental component. That they do pays tribute to their evolutionary provenance, and led some to question the involvement of instrumental responses at all (Mackintosh, 1983). The latter is implied, however, by the arbitrariness of the required avoidance actions in some experiments (although more arbitrary ones are slower to learn (Biederman et al., 1964;Ferrari et al., 1973;Hineline, 1977;Riess, 1971)).

Further, there is good evidence that the safety state that arises from successful avoidance acts as a Pavlovian aversive inhibitor. This is a state that predicts the absence of otherwise expected punishment. Importantly, as mentioned above, the values of aversive inhibitors at least partly share a common representation with those of appetitive excitators, as is demonstrated by their ability to affect subsequent learning in appetitive domains (a phenomenon known as transreinforcer blocking). That the safety state plays an important role in control is suggested by the fact that avoidance responses continue long after the Pavlovian aversive responses to the discriminative stimulus have extinguished (as they will of course do if avoidance is successful).

This places in the spotlight the role of the value attached to the warning stimulus(Bersh and Lambert, 1975;Biederman, 1968;De Villiers, 1974;Kamin et al., 1963;Mineka and Gino, 1980;Overmier et al., 1971b;Overmier et al., 1971a;Starr and Mineka, 1977). On one hand it has the Pavlovian power to initiate Pavlovian preparatory responses. It is also known to be able to suppress appetitive instrumental behaviour, in a similar fashion to conditioned suppression by an aversive Pavlovian predictor. On the other, it has the instrumental power to initiate an appropriate avoidance response.

The existence of a goal-directed component to avoidance is suggested by sensitivity to outcome in experiments that manipulate body temperature. Henderson and Graham trained rats to avoid a heat source when rats were themselves hot. When subsequently tested when the animals had been rendered cold, avoidance was attenuated, provided the rats had the opportunity to experience the heat source in their new, cold state (Henderson and Graham, 1979).

# 1.3 Neurobiology of aversive motivation and learning systems.

## 1.3.1 Ascending nociceptive pathways

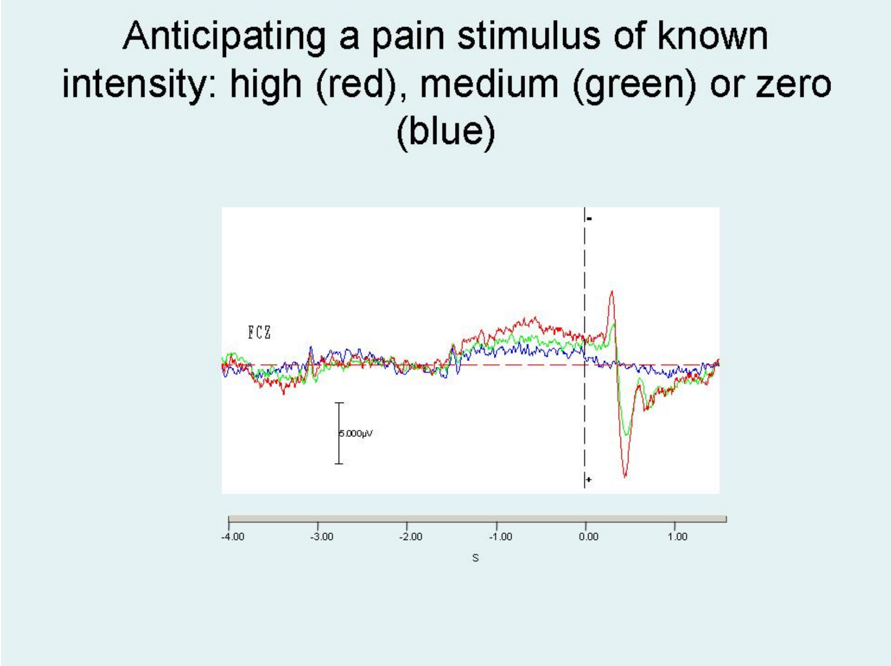
Injury comes in many different forms, in both routine life and scientific experiments. This diversity is reflected by the multitude of skin and tissue receptors which detect tissue damage (Hunt and Mantyh, 2001;Julius and Basbaum, 2001). This includes receptors for pressure, temperature (hot and cold), protons, inflammatory mediators, vascular damage, cell injury, etc. These receptors reside at the terminals of specific nociceptive neurons: – either the few large, fast, energy expensive myelinated A-delta fibres – typically responsible for acute, sharp pain; or the numerous (80% of all sensory neurons) smaller, slower, fibres responsible typically for long-lasting aching and burning pain. These nerves ascend the peripheral nerve to the spinal cord, have their cell bodies in the dorsal horn, and they synapse in certain specific layers (laminae) of the spinal cord (Craig, 2002). Nociceptive signals then ascend the spinal cord in two discrete pathways, the lamina 1,2 nociceptive-specific pathway and the lamina 5 wide-dynamic range pathway. As well as sending off branches to various brainstem nuclei, their main target is the thalamus, traditionally viewed as the gateway to the brain and cortex. In fact there are many other ways in which are likely to be important, for instance via the many brainstem nuclei. Beyond the thalamus, very many areas of the brain are involved in pain processing – including subcortical areas such as basal ganglia, amygdala, and hippocampus, and cortical areas such as somato-sensory, insula, orbitofrontal, and anterior cingulate (Jones et al., 1992). In fact extensive regions of the brain have been implicated in pain in some way, although it has been remarkably difficult to find any that are specific to pain. This fact makes it rather difficult to make ('reverse') inferences about function based on anatomy, a common fallacy in brain imaging research.

## 1.32 Pain anticipation

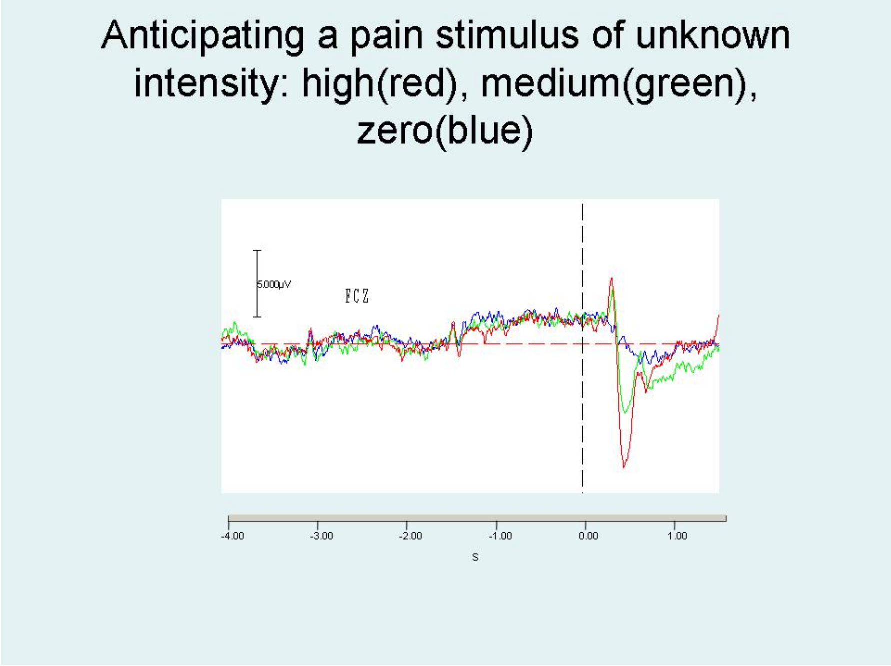
At the heart of neurobiological studies of the motivational basis of pain, embodied for instance by learning theoretic paradigms, is activity that occurs in anticipation of pain. This was first explored by Ploghaus and colleagues (Ploghaus et al., 1999), who used a simple (classical) conditioning design to look for anticipatory activity to thermal nociceptive stimuli, using fMRI. They found that activity in regions of anterior insula, anterior cingulate cortex, and medial prefrontal cortex correlated with the predictive period.

To look for basic representation of prediction related brain activity, related to the errors and anticipatory uncertainty predicted by theoretical accounts described above, we undertook a preliminary study, in which we studied the electroencephalographic activity of 15 subjects in a sequential pain prediction task. A series of auditory tones predicted the occurence of a painful laser stimulus to the right arm. The intensity of the stimulus was signalled before the auditory tones in half of the trials, whereas the other half were indicated as being uncertain.

We found a significant negative wave in the evoked potential in the time preceding the pain stimulus, that correlated with the predicted intensity of the subsequent pain (when it was known). This provides evidence that basic aversive value predictions can be detected in the brain following learning. Furthermore, when pain was predicted (in the ‘uncertain’ condition) and subsequently omitted, we found a significant negative wave following the omission. This activity may reflect a (negative) prediction error, evoked by the difference between expectation and outcome of the pain. The EEG characteristics and scalp topography of these activities is shown in figure 1.6, below.

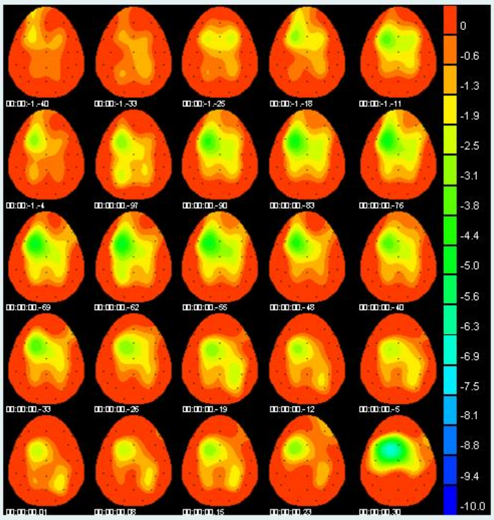


Anticipating a pain stimulus of known intensity: high (red), medium (green) or zero (blue)



Anticipating a pain stimulus of unknown intensity: high(red), medium(green), zero(blue)

Scalp topography, from 1400ms pre-stimulus to 300ms post stimulus (peak N2)



Expected (red) and unexpected (green) omission of pain stimulus at CPZ

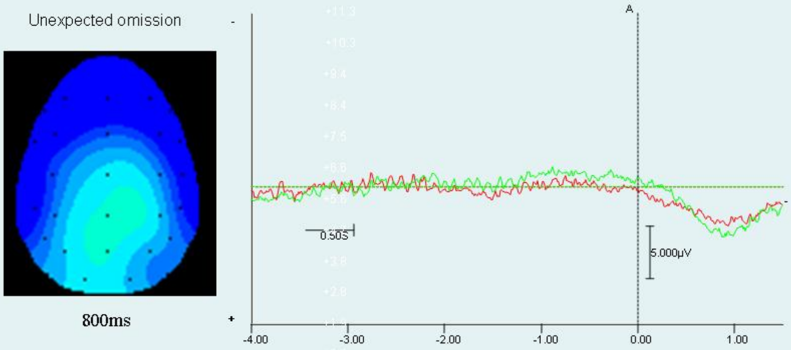


Figure 1.6. 15 subjects were studied with 32 channel ERP, with forearm CO2 laser-induced pain stimuli at 3 different intensities: High, Medium, Low (in fact zero - subjects told 'low' was likely to be imperceptible. The intensity was indicated 7 seconds pre-stimulus on a computer screen: on 50% of occasions, the forthcoming intensity was provided (i.e. the words 'high', 'medium' or 'low' printed on the screen).On the other 50% of occasions the forthcoming intensity information was withheld and 'unknown' printed on

the screen. The timing of the pain stimulus was indicated by a sequence of 3 countdown auditory tones (at 1.5 seconds interval). Throughout the experiment (both known and unknown expectations) subjects received 60% medium, 20% high and 20% low intensity stimuli in pseudo-random order.

Anticipation (panels 1-3): These data reveal a low frequency negative wave in the seconds prior to stimulus onset, maximal over FCz (. The amplitude of this negative wave correlated with the expected intensity (high: -2.68uV, medium: -1.74uV, low: -1.01uV). In the uncertain condition, the amplitude was comparable to the medium expected condition

Omitted stimulus potential (panel 4): In the unexpected compared to the expected low condition (omission), there was a significant late positive wave (corresponds to a negative prediction error), maximal over CPz.

This study provided pilot data for the subsequent PhD work, in the lab of Anthony Jones in Manchester.

These two findings – value and prediction error related activity, provide a basis for the subsequent experiments in this thesis, which use fMRI. The design of the ERP paradigm was further refined by Chris Brown, in Anthony Jones' lab, and studied using high density source localisation EEG. This found that the anticipatory activity correlated with activity in anterior insula (Brown et al., 2008a).

### 1.3.3 Aversive learning systems.

Existing studies of Pavlovian aversive learning implicate a network of predominantly subcortical regions that coordinate the acquisition of predictive value with the execution of responses. The amygdala is widely recognised as one of the principal brain structures associated with aversive Pavlovian learning (Gallagher and Chiba, 1996;H.Klüver and P.C.Bucy, 1939;LeDoux, 2000a;Maren and Quirk, 2004;Murray, 2007), especially in imaging neuroscience (Morris et al., 1998). Broadly, it consists of two functionally and anatomically distinct components, namely those that are affiliated with the central and basolateral nuclei. Both are heavily connected with extensive cortical and subcortical regions consistent with a capacity to influence diverse neural systems (Amaral and Price, 1984).

Early theories on the role of the amygdala centred on fear (WEISKRANTZ, 1956), in light of the key discovery that it acts as a critical seat of Pavlovian aversive conditioning (Maren, 2005;Quirk et al., 1995). More specifically, many elegant experiments have demonstrated that the basolateral amygdala, by way of its extensive afferent input from sensory cortical areas, is critical for forming cue-outcome associations, and that the central nucleus is critical for mediating conditioned responses, by way of its projections to mid-brain and brainstem autonomic and arousal centres (Kapp et al., 1992). In what became known as the 'serial model' of amygdala function, the basolateral amygdala is thought to learn associations, with direct projections to central amygdala engaging the latter to execute appropriate responses (LeDoux, 2000b).

In subsequent years, several key findings have emerged that have enriched this picture. First, the amygdala has been found to be critically involved in appetitive learning, in a similar way to its involvement in aversive learning (Baxter and Murray, 2002). Second, the central and basolateral nuclei often operate in parallel, as well as in series. This is thought to subserve dissociable components of learning, whereby the central nucleus mediates more general affective, preparatory conditioning, with the basolateral nuclei mediating more consummatory, value specific, conditioning (Balleine and Killcross, 2006;Cardinal et al., 2002). Third, rather than just executing Pavlovian responses, connections of both central and basolateral amygdala with other areas such as the striatum and prefrontal cortex are critical for integrating Pavlovian information with other decision making systems (Cardinal et al., 2002)

Single neuron recording studies have identified neurons that encode the *excitatory* Pavlovian value of rewards, punishments, as well as neurons that encode salient predictions independently of valence (Belova et al., 2007;Paton et al., 2006). In rodents, electrophysiological data implicate the amygdala in encoding appetitive inhibitors, suggesting that aversive value representations extend to opponent inhibitory stimuli (Belova et al., 2007;Konorski, 1967;Rogan et al., 2005;Seymour et al., 2005).

The amygdala is likely to mediate conditioned responses through connections with other brain regions such as the periaqueductal grey, hypothalamus, parabrachial nuclei, caudal pontine nuclei of the reticular formation, ventral striatum and ventral tegmental area (Fendt and Fanselow, 1999). Structures such the periaqueductal grey and anterior hypothalamus mediate primitive defensive, retaliatory and offensive responses, and encode essential motor patterning mechanisms for fighting (Adams, 2006). Other regions implicated in aversive value representations include the lateral orbitofrontal and anterior insula cortex(Calder et al., 2001;Craig, 2002;Jensen et al., 2006;Nitschke et al., 2006;O'Doherty et al., 2001;Paulus and Stein, 2006;Sarinopoulos et al., 2006;Seymour et al., 2005;Small et al., 2001), which, we note, are also interconnected with the ventral striatum (Mesulam and Mufson, 1982;Mufson et al., 1981).

Pavlovian appetitive learning also involves the amygdala, and indeed many responses, including preparatory arousal like responses, and specific consummatory responses, are mediated through connections to brainstem autonomic nuclei and hypothalamic centres (respectively). A substantial amount of research has also focused on the role of the ventral tegmental area, which sends dopaminergic projections to the ventral striatum, which underlies many aspects of Pavlovian appetitive responding. More elaborate value representations may rely on computations in the orbitofrontal cortex, notably those sensitive to manipulations of outcome value (for example, by altering motivational state, affective context, expected value, relative value, or counterfactual value(Baxter and Murray, 2002;Milad and Quirk, 2002;Nieuwenhuis et al., 2005;O'Doherty et al., 2001;Roesch and Olson, 2004;Rolls, 2000;Schultz, 2000;Sugrue et al., 2005;Tobler et al., 2005a;Tremblay and Schultz, 1999;Ursu and Carter, 2005)).

## 1.3 Reinforcement learning and computational neuroscience of aversive learning.

### 1.3.1 General principles of a computational framework

David Marr distinguished computational, algorithmic and implementational ‘levels’ of understanding systems neuroscience. At a computational level, one can specify the function that the system or structure under study evolved to perform – what, formally, is the problem that an animal must solve in a particular domain? At an algorithmic level, one can understand the mathematical strategy that the brain uses to solve or perform this function. And finally at an implementational level, one can address how this strategy is implemented in the various hardware of neurons and neural circuits in the brain. Although clearly these different levels reciprocally inform each other, and studying any one in isolation might be less profitable than appreciating the relationship between levels, a recognition of the fundamentally distinct nature of these levels provides a powerful and invaluable framework on which to study systems neuroscience (Marr, 1969;Marr, 1970;Marr, 1971).

Such an approach often exploits optimality principles, justified on evolutionary grounds (Todorov, 2004). There are many aspects of behaviour, particularly more complex cognitive processes (including human decision-making), where this may not hold, but the basic functions of aversive learning systems ought to be, in most environments, suitably primitive and evolutionarily conserved to permit reasonable hypotheses that assume optimal (or near optimal) processing.

The Marrian framework yields an approach to systems neuroscience in which mechanistic accounts of behaviour can be sought. Any model, psychological or otherwise, is specifiable mathematically: this does not constrain a model, it only forces an explicit description of the structure and parameters within that model, which are sometimes covert in traditional psychological models. The strength of this approach is that explicit predictions can be tested empirically and quantitatively. This renders them not only substantially less ambiguous, but also open for refutation. This ought to stimulate strongly hypothesis driven experiments, and provide a well-lit arena for different theories to be pit against each other.

Remarkably, pain and aversive learning, at least from a systems neuroscience perspective, has somewhat escaped a normative approach, in stark contrast to other related disciplines in affective (such as reward processing and decision making) and sensory (such as vision and audition) neuroscience. This is fortunate for pain neuroscience, since it allows the field to poach insights of these other disciplines. But this relationship may prove more symbiotic than parasitic, since methodological reasons mean that pain is sometimes a better modality to study general principles of behaviour (for example, the salience of pain lends itself better to studying higher order learning than less salient reward). As a corollary, one should also remain vigilant to the peculiarities of pain, that is, those features, and there may be many, that cannot be generalised across valences.

Aversive events share the core feature in the capacity to threaten to a lesser or greater extent the integrity and survival of the individual. This is fairly explicit in the case of pain, but a diversity of stimuli or events may be judged aversive, such as odours, tastes, loud noises, as well as social stimuli such as exclusion or reputation loss. From a behavioural viewpoint we can make a broad definition of aversiveness as the property describing things we would rather not have, or things we would do work to reduce or avoid. The somatic pain system provides an ideal system to study aversive motivation: it represents actual or imminent tissue injury, and from an introspective perspective pain is inherently and potently aversive.

## 1.32 Formalising motivation and learning

The central problem faced by any aversive motivational system can be approached by a body of theoretical and empirical research called Reinforcement

Learning. Reinforcement learning deals with problem of how an agent should optimise their behaviour in an unknown environment, through experience. At the heart of this approach are several key concepts. The first is that the agent has a representation of some quantity that specifies inherent preferences: positive events such as food, and negative events such as pain. If these events can be sensed, then the behavioural problem is one of maximisation (or minimisation in the case of punishments). Second is that the agent can learn about their environment through trial and error experience: choosing actions and observing the outcomes that are delivered. Knowledge of these outcomes can then be used to improve performance in the future.

This theoretical framework is common in many disciplines (such as economics, control theory and ethology) that aim to model how systems of any sort can learn about the environments they inhabit, and make decisions that maximize beneficial outcomes and minimize adverse ones (Camerer, 1995;Mangel and Clark, 1988;Puterman, 1994;Sutton and Barto, 1998). This framework is closely associated with dynamic programming (Bertsekas, 1995), and encompasses many different algorithmic approaches for acquiring information about an unknown environment, including learning from trial and error, and using that information to specify controls.

In typical natural cases of decision making, feedback for a choice is usually only available after some time has elapsed, and, potentially, also additional choices (as, for instance, in a maze). This problem of delayed feedback has played an important role in determining the nature of the neural controllers, with forms of prediction lying at their heart (Montague et al., 1996;Sutton and Barto, 1998). The essence of the solution to the problem of delayed feedback is prediction of the value of being in a particular situation (typically called a ‘state’) and/or performing a particular action at that state, in terms of the rewards and punishments that can be expected to accrue in the future. Different ways of making predictions underlie the different approaches to control.

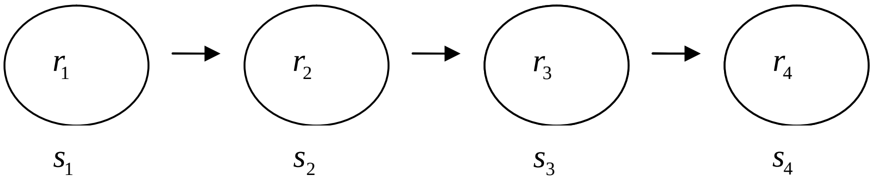
To specify the problem more formally, we can consider it a form of a general stochastic optimal control problem, defining:

- State S – the current combination of cues and context
- Action a – an action, moves agent from state to state
- Value V – the overall value of being in state S or of taking action a from state S
- Policy pi – determines which actions to take (e.g. always take the highest valued action)

We first consider the situation in which outcomes are delivered independently of any action taken. This describes the problem as purely one of prediction, as opposed to one of control, in which actions can be enacted that actually change the probability of an outcome. Prediction alone has a close parallel with Pavlovian conditioning.

The goal of prediction is to learn a value function, where the value represents the sum of future rewards or punishments expected to follow if the agent is in a particular state.

Consider the detrerminsitic sequence of state transitions below, where an agent moves from state s1 left to right, accruing reward r until the terminal state s4



The value of state s1 is the sum of the future expected reward from state s1:

$$V(s_1) = r_1 + r_2 + r_3 + r_4$$

If one assumes (exponential) discounting of future rewards, such that future rewards are considered less value than immediate rewards, then this becomes:

$$V(s_1) = \gamma r_1 + \gamma^2 r_2 + \gamma^3 r_3 + \gamma^4 r_4$$

Or more simply:

$$V(s_1) = \sum_{n=4}^{n=1} \gamma^n r_n$$

If one exploits the recursive relationship between successive states, then it can be seen that:

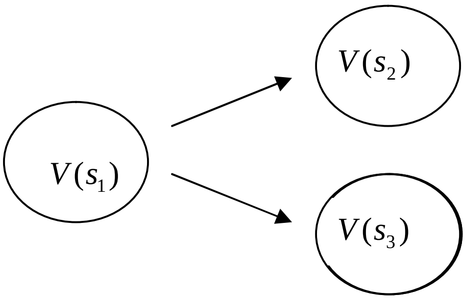
$$V(s_1) = r_1 + \sum_{n=4}^{n=2} \gamma^n r_n$$

Or alternatively:

$$V(s_1) = r_1 + \gamma V(s_2)$$

This simple equation specifies the relationship between the value of successive states. Reinforcement learning exploits this relationship, and uses value estimates to update the value of preceding states.

Now consider the more general situation in which the state transitions are not deterministic, but rather probabilistic. In this case, the value of being in a certain state is related to the value of future states, and weighted by the probability of reaching them. Consider the following state transition,



The value of s1 is equal to the sum of the product of the value of s2 and s3 and their state transition probabilties:

$$V(s_1) = r_1 + p(s_1 \rightarrow s_2)\gamma V(s_2) + p(s_1 \rightarrow s_3)\gamma V(s_3)$$

If actions a are permitted, which are defined by the agents policy pi, then with a Markov assumption (that dictates that the previous trajectory to a given state has no bearing on future state transition probabilties or rewards), then the general equation for the value of any state is given by the Bellman equation (Bellman, 1957):

$$
\begin{aligned}
&V^{\pi}(s) = E_{\pi}\{R_t \mid s_t = s\} \\
&= E_{\pi}\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s\} \\
&= E_{\pi}\{r_{t+1} + \gamma \sum_{k=0}^{\infty} \gamma^k r_{t+k+2} \mid s_t = s\} \\
&= \sum_{a} (s,a) \sum_{s'} \mathcal{P}_{ss'}^{a} [\mathcal{R}_{ss'}^{a} + \gamma E_{\pi}\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+2} \mid s_{t+1} = s'\}] \\
&= \sum_{a} (s,a) \sum_{s'} \mathcal{P}_{ss'}^{a} [\mathcal{R}_{ss'}^{a} + \gamma V^{\pi}(s')].
\end{aligned}
$$

Temporal difference learning provides a mechanism to learn the value function online. It exploits the recursive property of the Bellman equation to compare sequential value estimates, and uses a prediction ‘error’ to improve those estimates. The prediction error term is intuitive, and is equal to the numerical difference between the expected outcome from a particular state, and the subsequently experienced outcome when the next state is reached:

$$\delta = V(s_i) - (r_i + V(s_{i+1}))$$

The value of the preceding state is then updated according to the prediction error, and the learning rate $0 \le \alpha \le 1$:

$$V(s_i) = V(s_i) + \alpha\delta$$

The TD rule bears close similarity with the Rescorla-Wagner rule, the error-based algorithm from animal conditioning studies; and the delta rule (or Widrow-Hoff rule) in associative learning theory. The bootstrapping method (which describes learning between successive, sequential predictors) extends Rescorla-Wagner prediction errors to pure predictions themselves.

*Average reward prediction*

Another approach to acting in extended timeframes is to use average-reward TD learning, where one determines *relative* value as equal to the sum of future reward compared to the average reward rate.

$$\delta_{s(t)} = r_{s(t)} - \rho_t + v_{s(t+1)} - v_{s(t)}$$

$$\rho_{t+1} = \rho_t + \upsilon(r_t - \rho_t)$$

The average reward is slowly learned over time, with a learning rate much smaller than that for the phasic outcomes, ie. $\upsilon << \alpha$ .

*Action learning.*

Direct action learning can proceed in a very similar manner to state-value learning. Accordingly, the action (or ‘Q’) value is a quantity that reflects the amount of reward that can be expected after taking a certain action. This can be either the true expected value, or a preference weight, depending on exactly how one specifies learning. The Q values can be learned:

$$Q(t+1)_i \;= Q(t)_i \;+ \alpha \;\; \delta$$

$$\delta \;\;= r(t) - Q(t)_i \;- \rho(t)$$

using a prediction error as previously, and judged according to the average reward rate. This has parallels to Dayan’s advantage learning (Dayan and Abbott LF, 2001).

In chapter 6, we extend Q learning to deal with integrated appetitive and aversive components.

In summary, we present the hypothesis that the brain uses a temporal difference learning mechanism to learn about aversive events.

# Chapter 2: Methods.

The analysis of brain activity in awake, behaving humans has been studied for many decades using electroencephalography, which records electrical activity on the scalp with considerable temporal precision, but, despite new algorithms for source identification, somewhat less distinct anatomical localisation. The last 20yrs has seen two new revolutionary methodologies – Positron Emission Tomograpgy (PET), and functional magnetic resonance imaging (fMRI), which permit analysis of brain activity, inferred from regionally distinct increases in blood flow, with vastly improved anatomical precision, although considerably reduced temporal resolution, given the reliance on blood flow as opposed to directly assessing electrical neural activity.

fMRI allows inferences to made about simultaneous activity across the entire brain. It provides two basic sources of information: first, it provides spatial information allowing task-specific anatomical inferences, hence the commonly used term ‘functional brain mapping’, and second, it provides temporal information about the magnitude of task- specific brain responses, which allow, of particular interest here, assessment of dynamic changes in brain activity

The utility of fMRI rests on the basic and well-founded principles of functional localisation and specialisation. That is, macro-separable brain regions perform distinct physiological functions. This is supported by multiple lines of evidence, from the early brain stimulation studies of Olds and Milner [ref], to the reproducibility of specific cognitive deficits associated with particular neurological lesions, to neuro-physiological studies in primates in domains such as vision, where functional specialisation for colour, movement, form have provided spectacular evidence for the localisation of function.

Of particular interest from a learning perspective is the time course of activity, since emerging models of animal learning can be used to make rather specific predictions about both about the quantities which might be operationalised in certain learning processes (such as the prediction error), and how they should change through time. This might seem to imply that from the perspective of pure learning theory, simple anatomical localisation is rather uninteresting. However, in appropriately designed tasks, the simultaneous measurement of multiple and functionally distinct areas may allow disambiguation of learning related processes with multiple components, something far less straight-forward with conventional uni-dimensional physiological recording methods, traditionally employed by experimental psychologists, such as skin conductance, pupillary diameter, and heart rate measurement.

## 2.1 Physics of fMRI.

Magnetic resonance imaging relies on the electromagnetic properties of hydrogen atoms. The proton, which is positively charged, precesses on its own axis with a particular quantum magnetic spin, creating a very small electromagnetic field. Within a strong global electromagnetic field, as occurs within the bore of the MRI scanner, these spins will tend to appropriately align with the direction of the magnetic field, the field determining the precession frequency. The alignment of protons can be momentarily disturbed by applying brief radiofrequency pulses, which subsequently results in the release of a weak electromagnetic signal, detectable by the MRI scanner, as the protons return to their equilibrium state. Echoplanar imaging relies on the rapid provision of a spectrum of radiofrequencies that allow adequate sampling within time periods that fall inside that required to estimate the dynamics of the BOLD signal.

Functional magnetic resonance imaging provides an estimate of regional changes in blood flow. Haemogloblin consists of two distinct polypeptide chains, which are bound to an iron-rich protoporphyrin complex. Metabolically active tissue requires oxygen, which diffuses down a consistent concentration gradient from within the vasculature to the mitochondria, where it is used for oxidative metabolism to create ATP. Oxygen dissociates from haemogloblin in afferent capilliaries and becomes relatively deoxygentated in the efferent capilliaries and venules. Critically, oxygenated and deoxygentaed haemoglobolin have different magnetic properties, which alter the signal emitted by the protons of the hydrogen atoms within them. Oxygenated haemoglobin is diamagnetic, and so is little influenced by an external magnetic field, and consequently the phase coherence of proton spins. Deoxyhaemoglobin is paramagnetic which causes local magnetic field variations because of increased spin dephasing, because the four outer electrons of the Fe 2+ are now unpaired with oxygen. Blood-oxygen level dependent contrast (BOLD) exploits this natural difference in magnetic properties, and uses the contrast as an index of the oxygen uptake of peripheral tissues, which provides an indirect measure of tissue metabolism (Ogawa 1990, Turner 1991). Given that in the brain, the principle mechanism for variable tissue oxygen utilisation is neuronal activity, BOLD contrasts are proposed to offer a measure of neuronal activity. This is because of the relationship between regional cerebral blood flow and oxygen utilisation, given that vascular tone is under exquisitely sensitive control of local oxidative usage. Thus, increased energy demand from active neurons results in capilliary vasodilation. In fact, this vasodilation causes an effective over-shoot phenomneon, such that there is a relative *increase* in oxyhaemoglobin when metabolic activity increases. Since reduced deoxyhaemoglobin attentuates local susceptibilty effects, more active regions result in an increased signal intensity on T2 weighted images. BOLD images are, however, sucseptible to artefact due to large veins and arteries which may have more global variance due to uninteresting factors that influence cerebral blood flow.

## 2.2 Analysis of fMRI data.

Before fMRI images are analysed to assess significant task related effects within and between subjects, a number of pre-processing steps are performed, which represent predominantly spatial transformations for consistency of analysis across scans and subjects. fMRI analysis is voxel based, with each voxel typically being 2-3mm cubes. The spatial scale of the BOLD response, estimated by high-resolution optical imaging, is 2-5mm.

*Spatial realignment.*

Despite adequate head restraint in the fMRI, subjects may still move significant amounts during the course of an experiment, causing significant variance in the fMRI signal. To adjust for this, spatial realignment is performed to minimise scan-to-scan variance. Sequential scans are referred to the first scan, and the 6-parameters (in each 3 dimensional direction) are estimated for a rigid-body affine transformation that minimises the sum-of-squares difference between each. This transformation is applied using ‘sinc’ interpolation.

*Spatial normalisation.*

To allow comparison across subjects, scans are then normalised to a standard template schema. Here, we use the standard template of the Montreal Institute of Neurology (MNI), and all further references to anatomical co-ordinates are to this system unless otherwise stated. For this, the mean image of the re-aligned scans is taken, and the set of deformation parameters that maximises the likelihood of the data is found using an approximate iterative procedure (Gauss-Newton). These warping parameters can then be applied to all scans.

*Spatial smoothing.*

The are several reasons to spatially smooth data. First, by central limit theorem, Gaussian smoothing effectively makes error components more normal, thus strengthening the applicability of a parametric approach to the signal analysis. Second, smoothing can match the spatial scale of the data to the size of the effect anticipated (matched filter theorem), which will optimising efficiency in the detection of significant effects. Third, under random field theory, the metrics of the assumed underlying Gaussian field must be substantially larger than voxel size, which can be achieved by smoothing. Lastly, to accommodate functional anatomical difference between subjects, smoothing may well counteract the influence of inter-subject variability. Here, we generally adopt a smoothing kernel of 6-8mm given the predominant interest in subcortical structures such as midbrain and ventral striatum. However, kernels of 8-12mm are often superior for detecting population effects in cortical structures.

### *Statistical modelling*

In effect, the analysis up to this point provides a voxel-by-voxel time series of BOLD activation throughout the scanned volume of brain. The goal of the analysis is to relate in a statistically valid way, these time-series to some experimentally interesting manipulation. Thus, we want to make a statistical inference about regional brain activity given our experimental design.

The approach adopted, as is widely the case, is to propose a general linear model, and this the basis of Statistical Parametric Mapping (SPM, Wellcome Department if Imaging Neuroscience, London UK) used in this thesis. Thus, we apply standard parametric statistics to estimate voxel-wise statistical parameters in parallel. These parameters are typically T or F statistics, and their values displayed across the brain to identify regional effects.

SPM uses a mass uni-variate approach, and thus treats each voxel separately with respect to the experimental manipulation, and does not consider the covariance between voxel pairs as a multivariate approach would. Given the large number of voxels, the multivariate approach is highly inefficient, and under the security of an appropriate institution of Gaussian random field theory in protecting against the problems of multiple comparisons (see below) one can proceed to assay voxel-wise statistical parameters. In this thesis, the approach we take is based on classical inference, that is, we consider the evidence for the null hypothesis that some experimental manipulation has no effect on the signal in each voxel. Thus statistics are generated by estimating the size of an effect, its variance, and the error, in the data.

The general linear model assumes the generic form Y= beta * X + epsilon. Put linguistically, we propose that our observed data, Y, is a function of our experimental manipulation X, times a parameter beta that governs the size of the 'effect', and some residual error, or noise, or other effects (Friston 1995). Thus analysis is based on multiple linear regression, testing the null hypothesis that the estimated effect size of any individual regressor is zero. The central feature of the analysis thus becomes the design matrix – the temporal sequence of possible explanatory variables of the data. The design matrix will therefore include the particular effects that represent the manipulation that is proposed to modulate brain activity in some region, the so called 'effects of interest', plus any, and there may be many, other potential explanatory variables, which may be relatively uninteresting, often termed the 'effects of no-interest'. This may include things like session effects, uninteresting obligatory experimental manipulations, and even the movement parameters determined from realignment (above) to provide additional refinement of the model to account for variance not effectively removed by re-alignment procedures. Effects of interest may specifically relate to the influence of a single effect in specified direction, in which case one considers the effect size divided by its standard deviation, to give a T statistic, or by considering some (linear) combination of more than one effect by considering the relative variances, to compute an F statistic.

The design matrix considers the various effects that may influence our data, but we assume that most of these explanatory variables (ie, not effects like movement parameters) influence neural activity. However, our data represent the estimated blood flow, which is assumed to coupled to neural activity in some meaningful way. To incorporate this to our data, we typically apply prior knowledge about the nature of this coupling, namely, the shape of the haemodynamic response for some instantaneous burst of neural activity, into our statistical model. This is termed the haemodynamic response function (HRF), and in SPM is a synthetic, though biologically inspired (and validated), time-dependent vector, which peaks at about 5-6 seconds. With this in hand, we can estimate in what way our explanatory variables should influence our actual BOLD data, if they influence neural activity and the coupled blood flow in the manner proposed. Thus, in event-related designs, we effectively convolve (that is, multiply) the stimulus onset vectors in the design matrix with this synthetic function and use this as the regressors to which our multiple linear regression is applied.

Importantly, this is not the only way to make inferences, but it is more constrained. Outside these constraints, and if we are less sure about the nature of the haemodynamic response, one can institute a more flexible model. The most commonly used method of doing this is by proposing a set of (say, three) gamma functions, which form a basis set to which our brain response can be modelled. Less constrained still, we can use a set of small-duration (say, 2 second) rectangular impulse functions. However, the more basis functions we use, the less efficient, and less easy to interpret (not least because we have to estimate F and not T statistics) our results are.

One of the potentially serious hazards of the mass voxel-based univariate approach is the problem of false positives that arises from multiple comparisons. If each voxel was an independent observation, then the most appropriate method to correct for this is to perform a Bonferroni correction. However, voxels are not independent, and we can use the assumptions of random field theory to construct a more reasonable approach to this correction. Random field theory assumes that the error field conforms to a lattice approximation that has an underlying multivariate Gaussian structure, and secondly that these fields have a differentiable and invertible autocorrelation function.

The power of an anatomical inference grows with the precision of the prior hypothesis, and as such it is generally preferable to have some constrained hypothesis about the brain regions one expects to be involved in our experimental manipulation. This would come from a body of previous experimental work, which might include previous neuroimaging experiments, that allow us to pre-specify our region of interest, and then apply statistical correction derived from random field theory, within this region. This might ideally be an accurately shaped anatomical mask of a particular area, although practically, it is usually a sphere or 3D box centred on some pre-specified co-ordinate. In the absence of any prior anatomical hypothesis, one should ideally apply a whole brain level correction to the data.

## 2.3 Experimental design.

*Block and event-related designs.*

Consider experiments aimed to identify areas of the brain that respond to pain. Early fMRI and PET studies assessed the categorical effect of some experimental variable which was changed in different period – such as providing alternating periods of time in which a subject received thermal pain stimuli, and period of time in which they received non-painful thermal (warm) stimuli. The timecourse of presentations was typically analysed in a so-called box-car or block design – considering each period of activity as a whole, and making comparisons between them. However, the temporal precision of fMRI permits a more focused design, since the timecourse of the fMRI BOLD response allows disambiguation of individual stimuli. Thus, more recent designs can randomly alternate painful and non-painful stimuli and treat them individually, essentially as mini-boxcars of instantaneous duration (a so called delta, or stick function). These corresponds to an event-related design, and confers much greater flexibility in stimulus presentation, although may be less powerful if there is no anticipated (that is, cognitive) reason (such as habituation) why stimuli should not be presented in blocks.

*Design types.*

The simplest design types are subtraction designs. This rests on the proposition that the difference between two experimental tasks or conditions is the cognitive effect of interest. For instance, subtracting painful from non-painful conditions is proposed to identify areas of the brain specifically involved in pain, whereas subtracting highly painful from moderately painful conditions might be proposed to identify brain areas specifically involved in the processing the intensity of pain. However, subtraction designs can often be criticised because it is often possible to identify effects other than that of interest which are different between two conditions.

Considerably more powerful are parametric designs, which assess event-by-event differences in magnitude of a particular quantity. Thus, we might be interested in brain areas that are associated with subjective reports of pain intensity, so it is possible to linearly correlate subjective ratings with brain responses using parametric designs. Further still, this turns out to be a powerful tool when trying to identify brain responses that correlate with some potentially complex parameter. In learning experiments, one often has in mind a proposed computational model of how the brain might learn about some quantity – such as in reinforcement learning models of classical aversive conditioning experiments. These models might involve some key parameter that changes according to the complexities of the model in some determinable way, and modelling this can provide us with the predicted magnitude of this parameter through the course of our experiment. In this way, we can use fMRI to test the idea that such a signal exists, and then make the inference regarding brain activity in some (ideally, predicted) brain area is consistent with the predictions of our proposed model.

Multifactorial designs are essentially embedded subtractions, and allow assessment of how one experimental factor influence another. That is, they look at how the difference between tow levels of factor 1 is influenced by the difference in factor 2 – this would be termed a 2x2 factorial design, but of course potentially one can have multiple factors. The key advantage of factorial designs is that they can be used to assess interactions, on top of main effects.

*Population-level inferences.*

It is usually desired, and easily possible when we have potentially a sizeable subject population (in contrast with some rare patient groups, or monkey experiments), to make inferences that are generalizeable to the population. This requires an estimate of the variance between subjects and constitute random effects analysis, as opposed to an assessment of variance within subjects, constituting a fixed-effects analysis, in which we can only make our inference about that subject or particular group of subjects. Random effects analyses require taking some summary statistic to the group level, usually a contrast map from a within-subject analysis. The analysis is then usually classical, in which one tests the null hypothesis that the contrast map is zero.

## 2.4 Psychophysical measures:

*Pupillometry:*

Pupillometry is widely used to measure autonomic activity in experimental psychology, particularly in humans. The pupil is innervated jointly by sympathetic and parasympathetic afferents. There are two components to the pupillary response. The light reflex is the rapid constriction to a bright visual stimulus, and is attenuated with emotionally valanced cues. Subsequent to this, pupil approaches a new baseline level - which is greater for a broad range of emotional or arousing states. Analysis relies on classical statistical inference and is relatively well standardised (see Bitsios et al 2004). Blinks are removed by linearly interpolating across them. The data are baseline corrected (taking the mean diameter for 180ms before the cue). One can then take the peak minimum diameter (i.e. the amplitude) for the light response, and the mean diameter in the final 500ms before the delivery of the outcome.

*Reaction times***.**

Another useful index of emotional learning is by looking at task relevent influences on reaction times, particularly for tasks that are orthogonal to the manipulation of interest. Typically, reaction times are faster with emotionally arousing values, of either valence.

## 2.5 Pain stimulation.

There are a number of different experimental techniques available for stimulating ascending nociceptive pathways in humans, which have important differences in the physiological mechanisms they engender. Broadly speaking, different methods differ in the specificity with which they cause peripheral activation of c-fibres, a-delta fibres, and other non-nociceptive fibres. Electrical stimuli, widely used to elicit an painful state in experimental psychological investigations cause a relatively non-specific activation of predominantly aδ and aβ fibres. Current can be applied across two (oppositely charged) surface electrodes, or the electrodes can be needles placed subcutaneously or intra-muscularly. The use of electrical stimuli is often criticised by members of the pain community, particularly by those concerned with the anatomical differences between the spinothalamic (wide-dynamic range pathway, nociceptive specific pathway), and above, beause of this non-specificity, but this of much less concern to experimentalists concerned with learning theory, since as mentioned in chapter 1, the aversive qualities of a stimulus are dissociable form their sensory specific aspects. However, it might well be the case, and evidence has not been sought, that different types of pain may be more or less efficient, in experimental circumstances, at engaging aversive learning mechanisms. For instance, one might suppose that the fast aδpain might be better able to elicit conditioning in Pavlovian designs, since these are typically more efficient with short CS-US intervals. With a conduction velocity of less than 1m/s, c-fibre activation, for instance from the lower limb, may well take at least a second to arrive at the brain, which adds a non-trivial latency and, given the variance of conduction velocities of c-fibres, error component. In several experiments used in this thesis, we use a concentric surface electrode to deliver pain to subjects in the scanner.

This consists of central anode, which is pointed to allow good contact with skin, and a concentric circular cathode, which sits on the surface of the skin, and is typically placed on the dorsum of the hand. This electrode selectively activates aδ fibres, and was originally developed for use in studying facial pain. There is substantial subject to subject variation in the efficacy of this type of electrode since the ability to deliver current depends on the skin impedance, which varies widely according to subject skin character, body temperature, subject arousal and other factors, which become significant with limits on the maximum voltage utilisable. We note a critical safety issue with delivering electrical stimuli in the imaging environment, and that is that a rapidly altering electromagnetic field can induce large and potentially dangerous, currents. This often depends on intricacies of conductor topology, such that the diameter of a coil of wire. To protect against this, all electrodes used have high resistance (10 000 ohm) resistors placed within a few cm.

A number of other pain stimulation techniques are available. C02 or argon laser heat allows accurate and highly specific pain stimulation, but has several limitations: first it causes skin damage and the laser beam must be constantly moved around a significant area of skin to minimise burning, which is technically difficult in the fMRI environment. Second, it is logistically difficult, though not impossible to use in magnetic fields – the lasers themselves are ferromagnetic, so must be housed outside the scanner room, and the laser beam directed through shielded holes in the scanner room wall and directed to subject using a configuration of mirrors. Third, laser devices are currently very expensive.

Rapid, aδ fibre mediated mechanical pain can be delivered by pneumatic ballistic devices, which propel a ‘ball’ of pressurised air on to the surface of the skin. However, these devices are not available for use in the fMRI environment. Tonic pain can be delivered by the cold pressor test – inserting the periphery of a subject, such as the hand, into cold water. This causes an escalating cold thermal pain, and is a safe and reliable method for delivering tonic pain, but has the drawback that the pain is not constant, and rises progressively as the peripheral tissue gets colder. Alternative methods for delivering tonic pain include ascorbic acid injection, ischemia, intradermal or superficial application of mustard oil or capsaicin, and thermal heat. The latter two are used here and discussed below.

Thermal stimuli have become one of the preferred methods of pain delivery in fMRI and PET studies for several reasons. First, they are relatively pain specific, exciting thermal nocicpetive afferents on both aδ and c-fibres. Second, pain thresholds and tolerance are well studied, particularly in clinical neurophysiology, and known to be reliable and reproducible within subjects. Third, they can be used to deliver safe pain without causing skin damage through burning, and are susceptible to far habituation or sensitization than other methods. Fourth, there are now several commercially available fMRI compatible thermodes. These tend to be Peltier devices with water cooling facilities, which can deliver thermal stimuli in the typically used experimental range of up to 5 degrees per second, with heating typically being slightly faster than cooling.

In chapter 4, we aim to study relief of pain, in addition to pain itself, for which we use topical application of capsaicin, with overlying thermal stimulation, to elicit a state of tonic pain. Here, we use thermal cooling as a method of inducing a state of relief. This is not possible if thermal stimuli are used alone, because prolonged delivery of heat at sufficient levels to cause significant prolonged pain, such that relief is clearly and appetitively felt, will cause substantial skin damage due to burning. Capsaicin causes thermal hypersensitivity, allowing delivery of tonic thermal pain at much lower temperatures. Capsaicin, the ‘hot’ ingredient of chilli peppers, activates the TRP V1 family of receptors on peripheral nociceptive neurons, causing hyperalgesia. This provides an ecologically and clinically valid model of injury, since this process mimics the physiological changes that occur after many types of injury (such as burning).

# Chapter 3: Higher order aversive learning (expt. 1).

## 3.1 Introduction.

Predictions about potentially harmful stimuli should be available as early as they are reliable. In Pavlovian conditioning, chains of successively earlier predictors are studied in terms of higher order relationships, and have inspired computational theories such as temporal difference learning (Sutton RS and Barto AG, 1990). However, there is at present no adequate neurobiological account of how this learning occurs. Substantial evidence in humans and other animals has outlined a network of brain regions involved in the prediction of painful and aversive events (Buchel and Dolan, 2000;LeDoux, 1998;Ploghaus et al., 1999;Ploghaus et al., 2000). The majority of this work has concentrated on its simplest realization, namely first order Pavlovian fear conditioning. However, the predictions in this paradigm are rudimentary, revealing little of the complexities associated with sequences of predictors critical in psychological investigations of prognostication (Dickinson, 1980;Mackintosh, 1983). These latter studies led to a computational account called temporal difference (TD) learning (Sutton RS and Barto, 1990;Sutton and Barto, 1981), which has close links with methods for prediction (and also optimal action selection) in engineering (Sutton and Barto, 1998). When applied to first order appetitive conditioning, TD learning provides a compelling account of neurophysiological data, both with respect to the phasic activity of dopamine neurons in animal studies, and with BOLD activity of human functional neuroimaging studies (Friston et al., 1994;McClure et al., 2003;Montague et al., 1996;O'Doherty et al., 2003;Schultz et al., 1997;Suri and Schultz, 2001). However, the utility of TD models to describe learning beyond this simple paradigm remains largely unexplored. Here, we provide the first neurobiological investigation based on aversive and, importantly, sequential conditioning.

We used functional magnetic resonance imaging (fMRI) to investigate the pattern of brain responses in humans during a second order pain learning task. In brief, fourteen healthy subjects were shown two visual cues in succession, followed by a high or low intensity pain stimulus to the left hand (fig 3.1a).

Subjects were told that they were performing a reaction time study and were required to judge whether the cues appeared on the left or right of a display monitor. The second cue in each sequence was fully predictive of the strength of the subsequently experienced pain; however the first cue was only probabilistically predictive. Thus, on a small percentage of trials, the expectation evoked by the first cue would be reversed by the second. This allowed us to study the neural implementation of both the expectations themselves, and their reversals.

Two key aspects of most accounts of prediction learning are the predictions themselves (termed values), and errors in those predictions (Sutton and Barto, 1998). Fig. 1b shows the predictions (labelled TD value) associated with each trial type – these are calculated and revised as new stimuli are presented. Fig. 1c shows the associated prediction error. The success of TD learning in accounting for data on dopamine cell activity stems from the nature of this signal, which treats ongoing changes in predicted values on an exact par with actual affective outcomes. This prediction error signal drives learning by specifying how the predictions should change. In appetitive conditioning, the dopamine projection to the ventral striatum is believed to be a critical substrate for this signal, though apart from theoretical speculations about opponent processing (Daw et al., 2002), the equivalent for aversive conditioning is less clear. As in earlier work on appetitive conditioning, we used the TD model to generate regressors based on the values and prediction errors appropriate to each individual subject (O'Doherty et al., 2003). Statistical parametric mapping of the regression coefficients permits identification of regions associated with, and in receipt of information about predictions.

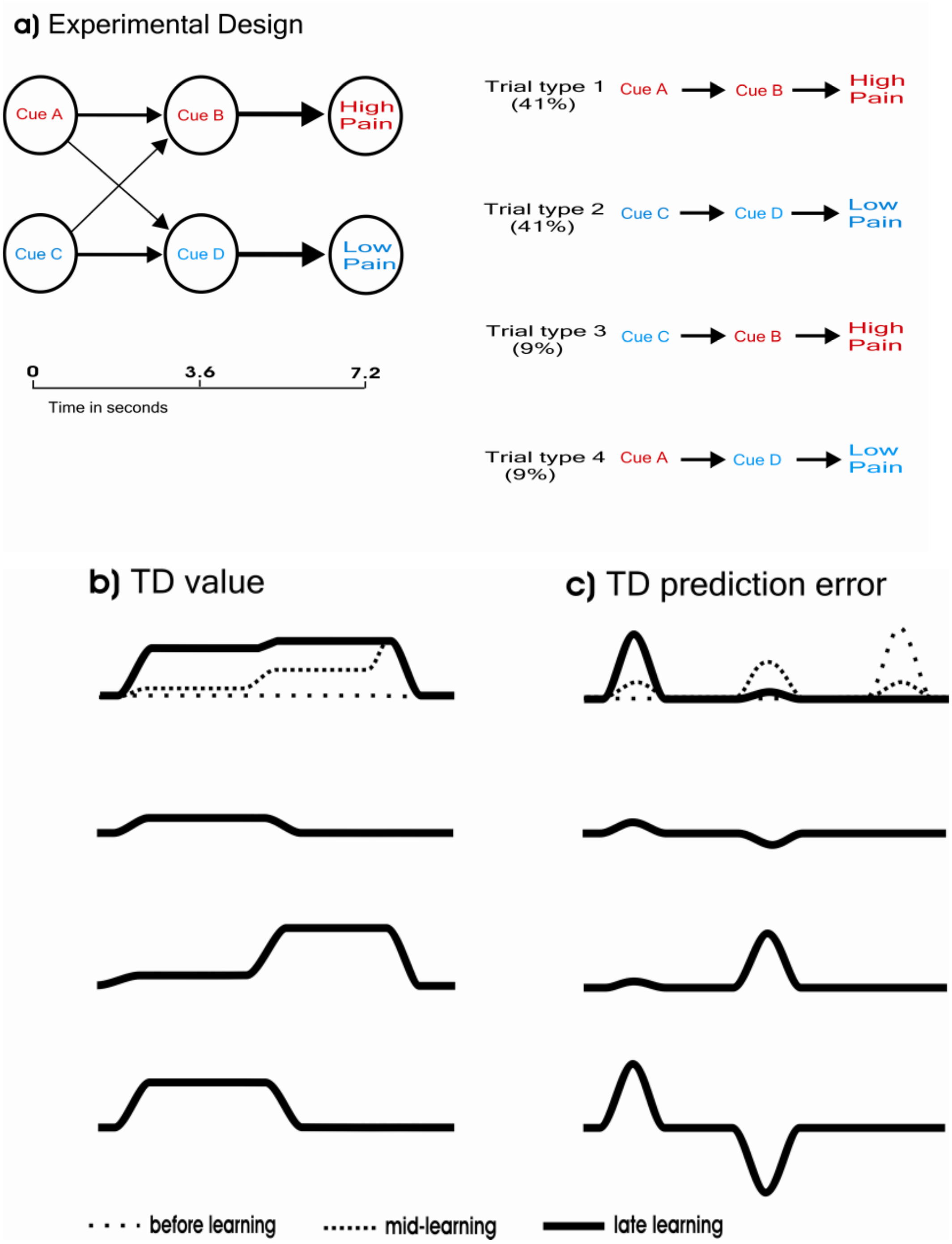

a) Experimental Design
Cue A
Cue B
High Pain
Cue C
Cue D
Low Pain
0
3.6
7.2
Time in seconds
Trial type 1 (41%) Cue A → Cue B → High Pain
Trial type 2 (41%) Cue C → Cue D → Low Pain
Trial type 3 (9%) Cue C → Cue B → High Pain
Trial type 4 (9%) Cue A → Cue D → Low Pain
b) TD value
c) TD prediction error
before learning
mid-learning
late learning

**Figure 3.1** Experimental design and TD model. **a)** The experimental design expressed as a Markov chain, yielding four separate trial types. **b)** TD model – Value: as learning proceeds, earlier cues learn to make accurate value predictions (i.e. weighted averages of the final expected pain). The 4 plots correspond to the 4 trials in a). **c)** TD model – Prediction error: during learning the prediction error is transferred to earlier cues as they acquire the ability to make predictions. In trial types 3 and 4, the substantial change in prediction elicits a large positive or negative prediction error. (Note: for clarity, early and mid learning are shown only for trial type 1).

## 3.2 Methods.

**Subjects.** Fourteen right handed volunteers participated in the study and gave informed consent. All subjects were pain free on the day of study, on no medication and had no history of neurological or psychiatric disease. The study was approved by the Joint Ethics Committee of the National Hospital for Neurology and Neurosurgery (UCLH NHS Trust) and Institute of Neurology (UCL).

**Stimuli.** We used an electrical pain stimulus delivered by a pair of silver chloride electrodes placed on the dorsum of the left hand 3cm apart. We delivered a 100Hz train of electrical pulses of 4ms pulse duration (square pulse waveform) for 1 second via an in-house built fMRI compatible electrical stimulator. Variation of current amplitude (0.5mA to 6.0mA) was used to deliver different intensity stimuli, set individually for each subject immediately before entering the scanner: subjects judged painfulness using a 10-point numerical rating scale (0 score = no pain, 1 point = just perceptible pain, 8 points maximum tolerable pain, 10 points = worst imaginable pain). We achieved mean intensity ratings of 2.9 for the low intensity stimulus and 8.0 for the high intensity stimulus. Post-hoc debriefing revealed no evidence of habituation or sensitization.

The visual cues were abstract coloured pictures of equal size and luminescence displayed on a screen projected into the scanner and visible by the subject through a mirror placed on top of the head coil. In each session there were four different cue stimuli, with a different set of pictures used in each, presented to the left or right and above or below the centre of the display screen. Pictures were fully counter-balanced across sessions and subjects.

Delivery of visual and electrical stimuli was controlled and synchronised with the MR scanner using Cogent 2000 software (Wellcome Department of Imaging Neuroscience, London, UK) implemented with Matlab 6.5 on a standard PC. The electrical stimulator was driven by a CED 1401 amplifier (Cambridge Electronic Devices, Cambridge, UK) with additional control using in-house software implemented on a separate PC.

**Task.** Subjects were required only to report the position of the cue (left or right) as quickly as possible, using a keypress with their right hand. No response was required to the pain stimulus and they were assured that the performance on the reaction time task bore no relationship to the intensity of subsequent pain. Subjects were not told the experiment was a learning task, and on post hoc debriefing no subjects were able to report the full set of cue – outcome contingencies.

**Experiment.** Each subject undertook two sessions. Each session represented a complete learning experiment, and consisted of 110 trials. Each trial consisted of the presentation of two cues in sequence followed by a pain stimulus. Each cue was presented for 3.6 seconds in immediate succession, and the offset of the second cue was followed immediately by the pain stimulus which lasted for one second. After this the next trial began after a variable (randomised) delay of 5+/-1.5 seconds.

Trials were divided into 4 types, labelled Types 1-4. Types 1 and 2 were the standard trial types and were presented at a frequency of 82% (41% each), and for a minimum of the first ten trials. In these trials the basic contingency of cue-cue-pain associations was set. Thus in trial type 1, cue A was followed by cue B which was followed by a high intensity stimulus; and in trial type 2 cue C was followed by cue D which was followed by a low intensity pain stimulus. Trials types 3 and 4 occurred randomly at a frequency of 18% (9% each type). In trial type 3, cue C was followed by cue B which was followed by high intensity pain stimulus, and in trial type 4, cue A was followed by cue D which was followed by a low intensity pain stimulus. This manipulation would be expected to induce second order TD prediction errors. The second cue always predicted the appropriate pain stimulus. The duration of each session was 13 minutes, after which the subjects underwent a high resolution structural brain scan.

**TD model.** We used a basic temporal difference learning model without eligibility traces or discount factor (TD(0))(Sutton and Barto, 1998) on a trial basis. Each trial was divided into three time points (first cue, second cue, pain stimulus). In our model, a state *s* is defined according to the particular stimulus present at that time (i.e. there are six states). Each state has a predictive value *V(s)* and a return *r* that represents the pain, with the high intensity pain stimulus assigned a return of one, and the low intensity stimulus and the visual cues zero. The predictive value of each state *V(s)* was initialised at zero. At each point in time *t* the prediction error *δ* is defined as

$$\delta = r + V(s_t) - V(s_{t-1})$$

In effect, this is the difference between successive value predictions taking into account the currently observed return. The previous value predictions are then updated according to the algorithm

$$V(s_{t-1}) = V(s_{t-1}) + \alpha\delta$$

where α is the learning rate. We used a learning rate of α=0.5 (see analysis below). The sequence of cue and pain stimuli for each subject entered this basic computational model to produce the prediction error and value at each time point throughout the entire session. These were then used as parametric regressors to analyse the imaging data.

**Data acquisition and analysis.** We acquired T2*-weighted EPI images with blood oxygen-level dependent (BOLD) contrast on a 1.5T Siemens Sonata magnetic resonance scanner. To optimise signal recovery in basal forebrain and midbrain structures, we used a tilted plane acquisition sequence (30 degrees to the AC-PC line, rostral > caudal) designed to minimise signal dropout due to susceptibility artefact(O'Doherty et al., 2003) and performed z-shimming in the slice-selection direction. Imaging parameters were: echo time 50ms, field-of-view 192mm, in-plane resolution 3mm, slice thickness 2mm, interslice gap 1mm. We acquired 280 volumes plus 5 dummy scans for each session, with a repetition time of 3.6 seconds. High resolution T1-weighted structural images were acquired after the experiment and co-registered with the mean EPI images, and averaged across the 14 subjects to allow group level anatomical localisation.

The images were analysed using SPM2 (Wellcome Department of Imaging Neuroscience, London UK). Functional scans were pre-processed by spatial realignment, normalisation to a standard EPI template, and spatial smoothing with an 8mm (full-width, half maximum) Gaussian kernel. The images were then analysed in an event-related manner, with the events defined by the onsets of all stimuli encoded as delta functions. To construct the regressor for the basic TD analysis, we multiplied the delta function with the TD prediction error at each event, provided by the computational model for each subject, and convolved the ensuing stimulus function with a canonical haemodynamic response function (HRF). To emulate a random effects analysis the parameter estimates (i.e. the regression coefficients) were taken to a second level group analysis using a one-way ANOVA.

The group level SPMs were initially thresholded at $P<0.001$ uncorrected (as displayed in figs 2 and 4). To correct for multiple comparisons, we used small volume corrections in our areas of interest, based on data from previous investigations of Pavlovian aversive conditioning from our laboratory (Buchel et al., 1999). Specifically, we used coordinates of ventral putamen (Right: 24,6,-6. Left: -18,9,-6), anterior insula (Right: 48,12,-6. Left: -54,12,-9), anterior cingulate (0,27,18), right amygdala (24,-3,-24) and cerebellum (Right: 24,-60,-30. Left: -30,-51,-30 from ref 14). We defined areas in substantia nigra, upper brainstem and dorsal striatum based on the anatomy from our mean structural image. Small volume corrections were 8mm radius spherical volumes. We report significant regressions using a family-wise error correction at $p<0.05$.

To explore the influence of the TD learning rate parameter, we numerically calculated a first order derivative of the prediction error with respect to the learning rate around a value of $\alpha=0.5$ for each subject. This was to linearise the prediction error with respect to the learning rate by approximating a Taylor expansion, allowing us to add the derivative as a separate regressor. SPMs of the appropriate F test failed to reveal a substantial effect of variation in $\alpha$ from 0.5, suggesting a learning rate of 0.5 to be approximately optimal. In support of this conclusion, learning rates of 0.2 and 0.8 gave similar though less robust responses.

To characterise the impulse responses in the right ventral putamen and anterior insula cortex (fig 3 and fig 4b), we performed a supplementary analysis using a flexible basis set of 2 second duration finite impulse responses for each of the four trial types. Within this design, each trial was treated as an event with the onset being the time of the first cue for each trial. We removed the first 10 trials from this analysis, during which early learning was taking place. On a subject by subject basis, we took the peak voxel from the original TD analysis in the area of interest, and plotted the time course in terms of the estimated impulse response to each trial type. These were then averaged across sessions, and subjects.

For the analysis of value, we used the sum of TD predictive value and the pain (return) value (given that our design is not optimal for distinction of the two) and treated the prediction error and pain as effects of no interest. To ensure reporting of purely predictive areas, we applied a mask (at $p < 0.05$, uncorrected) of areas showing significant differences in activity in the cue periods from the finite impulse response analysis.

To provide a behavioural index of conditioning, we took the mean reaction time to the first cue in each trial (i.e. cue A and Cue C, fig1) in the final third of each session (i.e. when conditioning ought to be robust). This was averaged across the two sessions for each subject, and then taken to a second level group analysis using a two-tailed t-test.

## 3.3 Results

Conditioning was demonstrated by a significant difference in reaction time to the first cue according to trial type after learning (high pain predicting cue 637ms; low pain predicting cue 616ms, $p<0.05$ two-tailed t-test).

In the analysis of fMRI data, the prediction error was highly correlated with activity in bilateral ventral putamen, right head of caudate and left substantia nigra (fig 3.2).

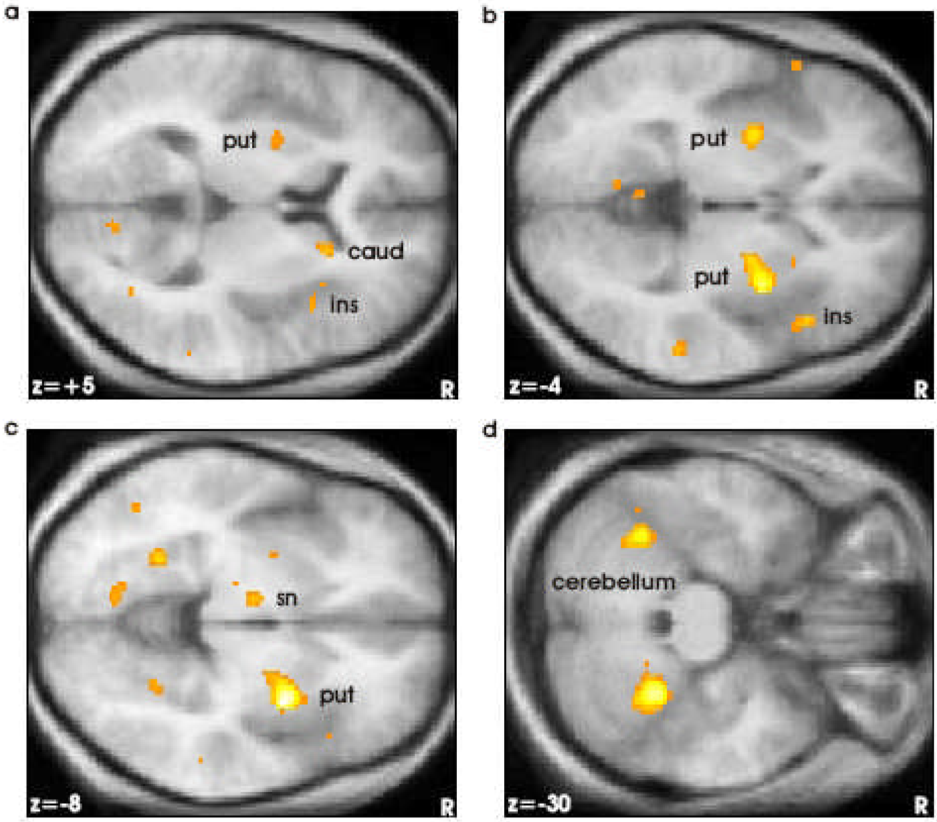


Figure 3.2 TD prediction error - statistical parametric maps. Areas showing significant correlation with the TD prediction error. Peak activations (MNI coordinates and statistical z scores) are right ventral putamen (put; (32,0,-8); z = 5.38), left ventral putamen (put; (-30,-2,-4); z = 3.93), right head of caudate (caud: (18,20,6); z = 3.75) left substantia nigra (sn: (-10,-10,-8); z = 3.52), right anterior insula (ins; (46,22,-4); z = 3.71), right cerebellum ((28,-46,-30); z = 4.91), left cerebellum ((-34,-52,-28); z = 4.42).

Correlations were also noted bilaterally in cerebellum and right insula cortex. Fig.3.3 shows the estimated responses in the right ventral putamen. As the most straightforward model coupling prediction error to BOLD signal might predict, positive (a: trial types 3 minus 2) and negative (b: 4 minus 1) prediction errors at various times in the trial are clearly represented, as is the biphasic form of the prediction error in trial type 4 (c: contrasted with type 2).

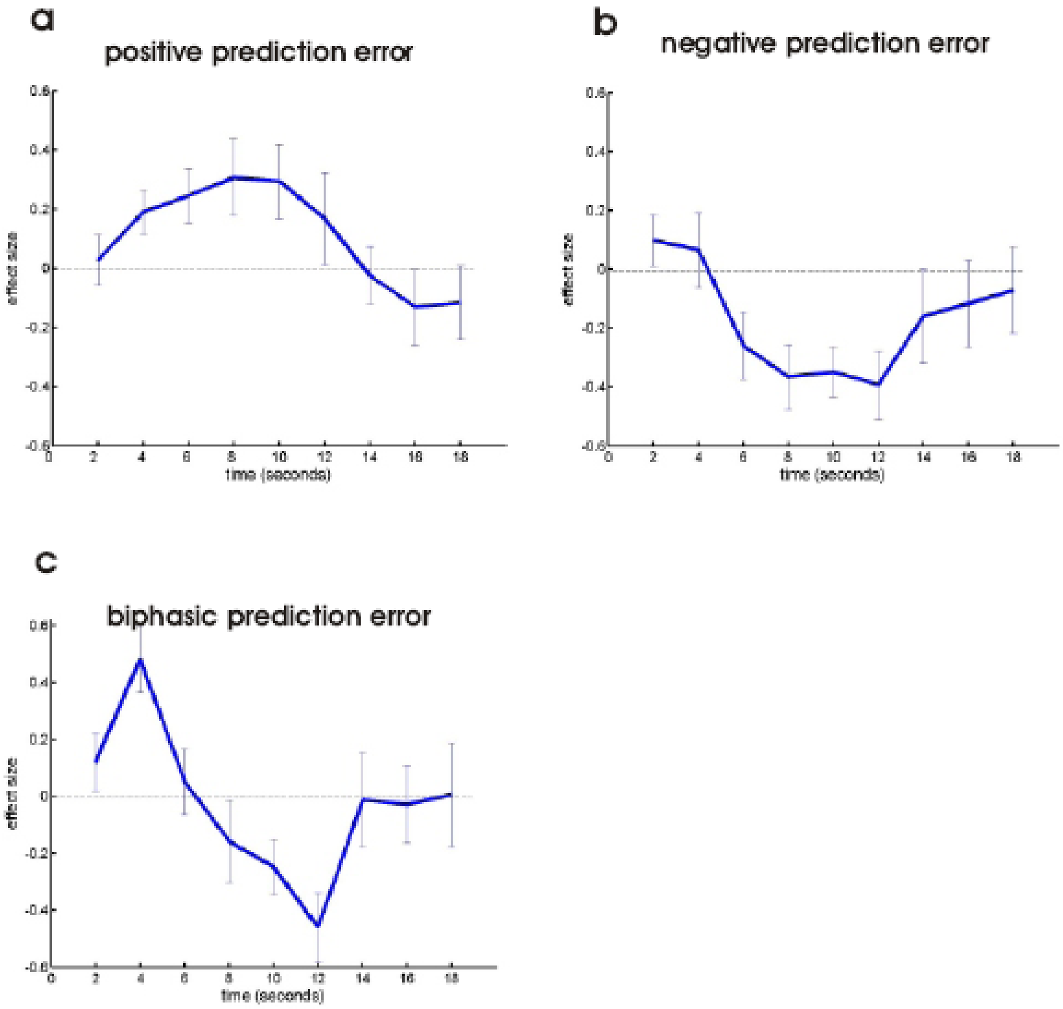


**Figure 3.3** TD prediction error – impulse responses. Time course of impulse response to higher order prediction error in right ventral putamen: **a)** positive prediction error (contrast of trial type 3 and 2), **b)** negative prediction error (contrast of trial type 4 and 1), **c)** biphasic prediction error: positive at the first cue becoming negative at the second (contrast of trial type 4 and 2).

We also investigated the representation of value (for reasons of analysis, combining the predicted and experienced value) by including the value term in our regression model. This revealed correlated activity in right anterior insula cortex (fig 3.4a). The estimated response is illustrated in figure 3.4(b). In addition, we found value-related responses in brainstem (fig 3.4a). Precise anatomical localisation of brainstem activation is difficult with standard neuroimaging, though we note the consistency with the likely location of dorsal raphe nucleus. We also observed value-related responses in anterior cingulate cortex and right amygdala, which did not survive statistical correction for multiple comparisons.

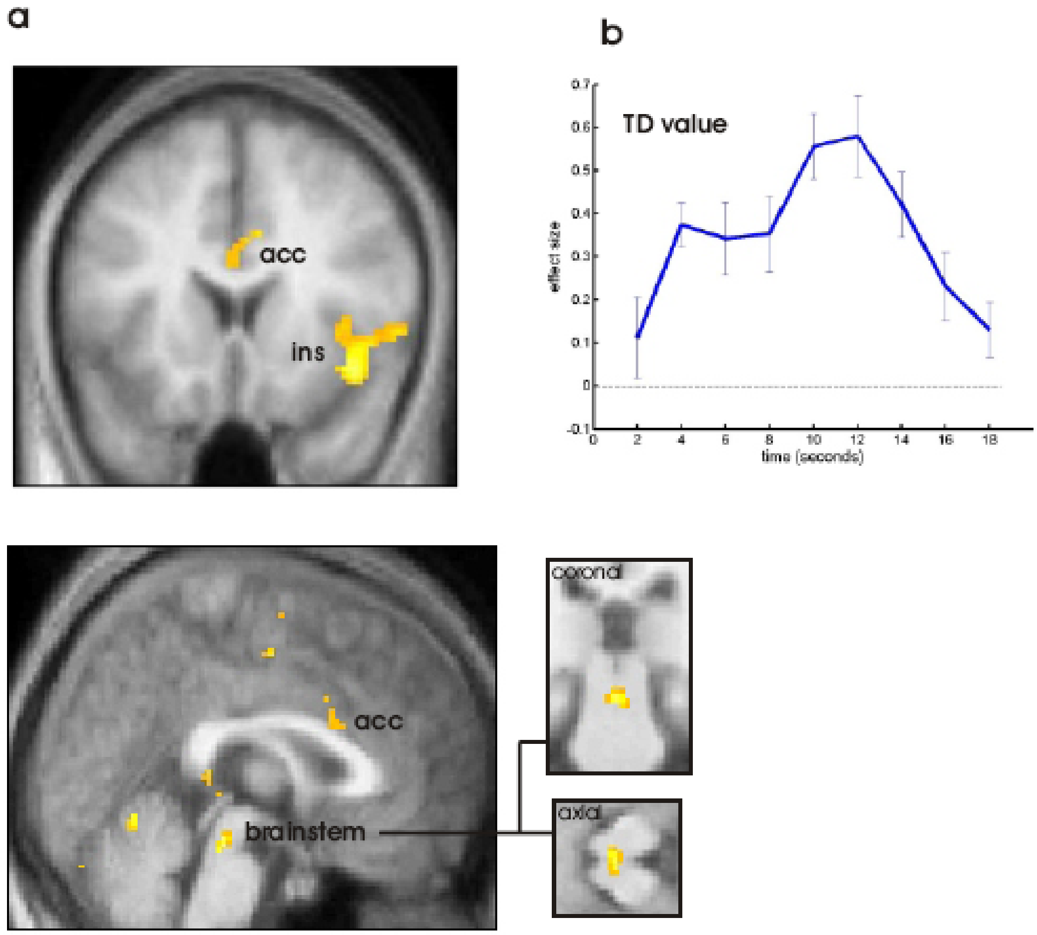


**Figure 3.4** TD value – statistical parametric maps and impulse response in right anterior insula. **a-b)** Areas showing significant correlation with the TD value. Peak activations (MNI coordinates and statistical z scores) in right anterior insula (ins; (42,16,-14); z = 4.16), brainstem ((0,-28,-18); z = 3.89) and anterior cingulate cortex (acc; (8,12,32); z = 3.82). Coronal and axial slices of brainstem activation are shown, highlighting localisation to dorsal raphé nucleus. **c)** Time course of impulse response in right anterior insula cortex, from contrast of trial types 1 and 2.

## 3.4 Discussion

The striking resemblance between the BOLD signal in the ventral putamen and the TD prediction error (fig. 3.3) offers powerful support for TD. This is particularly the case in a second order paradigm, since this captures the cue-to-cue bootstrapping of value predictions that lies at the heart of sequential prediction methods.  Other dynamic models of Pavlovian conditioning, such as the SOP models, do not involve this signal (Brandon et al., 2003), and deal instead with predictions that are immediately tied to outcomes. That is, they don’t learn using value estimates, but require actual outcomes. Our result adds to the growing body of neural and psychological data supporting the biological basis of TD theory. In a framework called the actor-critic model for instrumental conditioning (and some variants) (Barto AG et al., 1983;Barto et al., 1990), the same prediction error signal is also used to train stimulus-response habits (called policies), ultimately leading to choice of best possible actions (Barto, 1995). Again, this has been much more intensively studied from the perspective of appetitive than aversive conditioning. Importantly, the higher order process demonstrated here is a crucial substrate for learning in changing and uncertain conditions that characterise real environments, and in principle is capable of supporting complex behaviours.

Our findings add to the existing pharmacological, electrophysiological, functional imaging and clinical evidence regarding the involvement of the striatum in aversive processing and learning (Chudler and Dong, 1995;Schoenbaum and Setlow, 2003; Levita et al, 2002). Given that the BOLD signal in the same region is correlated with temporal difference prediction errors for rewards (McClure et al., 2003;O'Doherty et al., 2003), this structure may hold the key to understanding precisely how aversive and appetitive information are integrated to lead to motivationally appropriate behaviour in the light of (predictions of) both.

At present, the nature of the phasic aversive prediction error signal is not clear. Substantial psychological data suggest the existence of separate appetitive and aversive motivational systems that act as mutual opponents over a variety of timecourses (Dickinson and Dearing MF, 1979;Grossberg, 2000;Solomon and Corbit, 1974). Given the (not unchallenged) suggestions that dopamine neurons in the ventral tegmental area and substantia nigra report appetitive prediction error, it has been suggested that dorsal raphé serotoninergic neurons may encode aversive prediction error (Daw et al., 2002). It is of interest that we show prediction (value) related responses in an area that incorporates this nucleus. There is an active debate about the involvement of dopamine in aversive conditioning (Mirenowicz and Schultz, 1996;Romo and Schultz, 1989), and an alternative possibility is that dopamine reports both aversive and appetitive prediction errors.

Our findings have important implications for our understanding of human pain. Existing imaging studies have concentrated more on phenomenological aspects of pain processing. Here we have specifically explored aspects of the function of pain. Notably, substantial evidence indicates that the experience of pain is modified by prior conditioning (Ploghaus et al., 2003). Here, we demonstrate regionally distinct neuronal responses that are consistent with established computational processes that provide a mechanism through which the affective and motivational aspects of pain can be modulated.

# Chapter 4. Appetitive and aversive Pavlovian learning of phasic relief and exacerbations of tonic pain (expt. 2).

## 4.1 Introduction

Self-preservation and evolution ordain that animals act optimally or near-optimally to minimise harm. One of the principal mechanisms for detecting harm is the pain system, and early prediction is essential to direct appropriate pre-emptive behaviour. However, any simple correspondence between predicted sensory input and behavioural output is challenged by considering the nature of relief: for example, mild pain will be rewarding if it directly follows severe pain. This illustrates a critical issue in our understanding of pain relief as an affective and motivational state (Cabanac, 1971;Craig, 2003;Fields, 2004), and poses a broader question in emotion research: how do the neural processes that underlie motivation adapt to the context provided by the ongoing affective state?

According to psychological theories (Grossberg, 1984;Konorski, 1967;Schull, 1979;Solomon and Corbit, 1974), tonic aversive states recruit reward processes to help direct behaviour toward homeostatic equilibrium (which becomes the motivational goal). This may offer insight into why relief is often pleasurable, for example, the experience of cooling oneself in a swimming pool on a hot day. Indeed, the euphoria of relief has been used to help explain a number of seemingly paradoxical behaviours from sky-diving to sauna-bathing (Solomon, 1980b), in which relief is thought to become the dominant motivational drive. Despite supportive psychological evidence (Daw et al., 2002;Dickinson and Dearing MF, 1979;Solomon, 1980a;Tanimoto et al., 2004) direct observations of neural activity consistent with such appetitive processes are lacking.

Conceptually related issues arise in diverse areas such as engineering, economics and computer science, and offer potential insight into the underlying neural processes involved in relief in animals. Notably, reinforcement learning models have proved particularly useful in formalising how the brain learns to predict rewards and punishments (Barto, 1995;Dayan and Balleine, 2002;Montague et al., 1996;O'Doherty et al., 2003;Schultz et al., 1997;Seymour et al., 2004;Sutton and Barto, 1998). These models learn to make predictions by assessing previous contingencies between environmental cues and motivationally salient outcomes. In theory these models can be extended to deal with tonic reinforcement and relief, by computing predictions relative to an average rate of reinforcement, rather than according to absolute values (Mahadevan, 1996;Schwartz, 1993). However, the extent to which average reward reinforcement learning strategies are implemented in the brain is still unclear. With respect to pain, this may have added importance since motivational predictions (of pain or relief) are thought to exert substantial influence on the subsequent perception of pain (Fields, 2000;Price, 1999). Understanding the neural mechanisms by which predictions are learned is therefore a key component to our understanding of how the brain intrinsically modulates pain in physiological and clinical situations.

We used fMRI to investigate the pattern of brain responses in nineteen healthy subjects as they learned to predict the occurrence of phasic relief from, or exacerbations of, tonic pain (see methods). We employed a first order Pavlovian conditioning procedure with a partial (50%) reinforcement schedule (figure 4.1). Tonic pain was induced using the capsaicin-heat model. Capsaicin is the pain-inducing component of chilli pepper, and induces sensitisation to heat by activation of temperature-dependent TRPV1 ion channels expressed on peripheral nociceptive neurons. This temperature sensitivity allowed us to deliver constant but easily modifiable levels of pain for long durations, adapted for each individual subject, at temperatures which do not cause skin damage. This provides a unique experimental tool to study pain, since it specifically permits investigation of the neural processes underlying the *offset* of pain – that is, relief. The model has the further advantage that it induces the characteristic molecular and cellular changes that mimic physiological injury, and so presents a biologically realistic model of relief in natural and clinical environments.

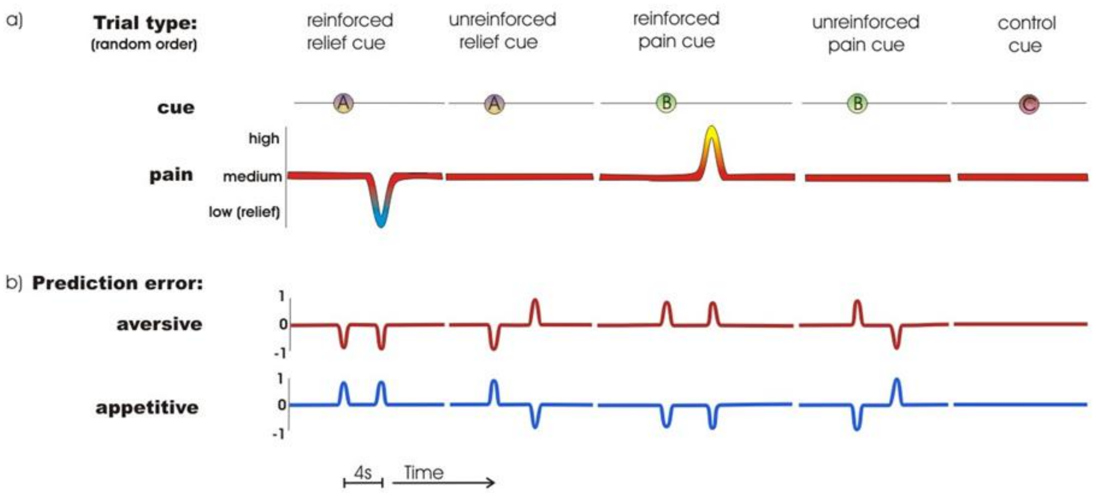


**Figure 4.1** **a) Experimental design.** There were five trial types: Cue A was followed by a temperature/pain decrease on 50% of occasions (reinforced and un-reinforced relief cue), cue B was followed by a temperature/pain increase on 50% of occasions (reinforced and un-reinforced pain cue), and cue C was followed by no change in temperature/pain (control cue). **b) Appetitive computational model – predicted neuronal response.** Schematic showing the mean representation of the temporal difference prediction error according to the different cue types, where relief is represented as reward. **c) Aversive computational model – predicted neuronal response.** Schematic showing the aversive temporal difference prediction error, which treats pain exacerbation as punishment. Note b) and c) represent the average predicted *neuronal* response: the corresponding predicted BOLD response is shown in figures 3c and 4c, respectively, following convolution with a canonical haemodynamic response function

We applied capsaicin topically to an area (12.5cm$^2$) of skin on the left leg causing a localised area of burning pain (which feels similar to sunburn), and manipulated the intensity of this pain with an overlying temperature thermode that matched the capsaicin patch. Temperature was adjusted for individual subjects to aim for evoking an average baseline magnitude of pain rated as 6 on a 0 to 10 categorical scale. Phasic decreases in the baseline temperature to 20°C caused complete relief of pain, and temperature increases caused exacerbation. We used visual cues (which were abstract coloured images) as Pavlovian conditioned predictors of these changes. Thus, in the fMRI scanner, subjects learned that certain images tended to predict imminent relief or exacerbation of pain.

We used a reinforcement learning (temporal difference) model to identify neural activity consistent with reward-like processing. The characteristic teaching signal of these models is the prediction error, which is used to direct acquisition and refinement of expectations relating to individual cues. The prediction error records any change in expected affective outcome, and thus occurs whenever predictions are generated, updated or refuted. By treating relief of pain as reward, and exacerbation as negative reward, we sought to identify activity that correlated with this prediction error signal. We calculated the value of the prediction error for each subject, according to the sequence of stimuli they received, to provide a statistical predictor of fMRI data (as has been done previously (O'Doherty et al., 2003;Seymour et al., 2004;Tanaka et al., 2004)). The use of a partial (probabilistic) reinforcement strategy, in which the cues are only fifty percent predictive of their outcomes, ensures constant learning and updating of expectations, and generates both positive and negative prediction errors throughout the course of the experiment (Figure 4.1b). Thus, inference is based on identification of this dynamic and highly characteristic signal.

## 4.2 Methods:

*Subjects:* 33 healthy right handed subjects (14 in a behavioural version of the task, and 19 in the fMRI version of the task), free of pain or medication, gave informed consent and participated in the study, approved by the Joint National Hospital for Neurology and Neurosurgery (UCLH NHS trust) and Institute of Neurology (UCL) Ethics Committee. Subjects were remunerated for their inconvenience (40GBP).

*Stimuli*: *Capsaicin model.* We applied topical 1% capsaicin (8-methyl-N-vanillyl-6-nonenamide, 98%, Sigma-Aldrich, Gillingham, UK, diluted in 5% ethanol-KY jelly) to the lateral aspect of the left leg over an area of 2.5x5cm, under an occlusive dressing, and left for 40 minutes, after which all subjects reported feeling persistent (though bearable) pain, at which time the capsaicin and dressing was removed and the skin cleaned. A thermode matching the size of the capsaicin application area was applied with a loose tourniquet (easily removable in case of unbearable pain) to the treated skin. Temperature was then manipulated using an fMRI compatible Peltier thermode (MSA thermotest, Somedic, Sweden). Phasic variations in temperature were achieved at a rate of 5°C/sec, to the predetermined upper and lower levels, and controlled by in-house designed software.

*Pre-experimental set-up*: Before the experiment, required temperature levels for each individual subject were set by slowly increasing the cutaneous temperature overlying the capsaicin treatment site from 20°C in 0.5°C steps, with continual monitoring of pain ratings (on a 0-10 rating scale), to achieve a baseline level of 6/10. Subsequently, subjects received progressively higher phasic increases to determine a satisfactory temperature for the pain exacerbations, to at least 8/10 (just-tolerable). Pain relief was induced by phasic cooling to 20°C, which abolished pain in all subjects.

We obtained subjective ratings of pain for the increase, baseline and decreases in pain. We asked the subjects, ‘Can you give a score, on a scale of zero to ten, as to how painful the pain is, where zero is no pain at all, and 10 is the worst imaginable pain’. We also took subjective ratings of pleasantness for the phasic relief. We first asked the subjects ‘Did you find the change in temperature unpleasant or pleasant’, to check that no subjects found the cooling as unpleasant, and then ‘Can you give a score, on a scale of zero to ten, as to how pleasant you found it, where zero is not at all, and ten is highest imaginable pleasure’. Phasic changes were repeated with pain and pleasantness ratings on capsaicin treated skin and on distant area of non-capsaicin treated skin on the same limb well beyond the area of secondary hyperalgesia, and repeated at the end of the experiment. We achieved mean ratings (standard error in parentheses) for the baseline tonic pain of 5.5/10 (1.1) on capsaicin treated skin and 0.9/10 (1.5) on untreated skin. Phasic increases were rated at 9.3/10 (0.9) for capsaicin treated skin and 3.3/10 (3.6) on untreated skin. Phasic decreases (relief) were rated at 7.0/10 (2.4) (pleasantness scale) and 4.6/10 (2.3) on untreated skin. All comparisons (treated vs untreated) were significant at P<0.01 with corresponding t-tests. Following transfer into the scanner (or behavioural testing) room (with the thermode attached) subjects were in pain for approximately 40mins to 1 hour by the time the experiment started.

*Stimuli: Visual cues.* The visual cues were abstract coloured pictures.

*Task*: The task was a classical (Pavlovian) delay conditioning paradigm of temperature increases (exacerbations of pain) or decreases (relief of pain). Visual cues were presented for 4 seconds, at the end of which the phasic pain perturbation was applied, for 5 seconds. The precise timing was determined in psychophysical pilot testing (to accommodate thermode and C-fibre latencies). There were three different visual cues, each presented 30 times. Cue A (relief related cue) was followed by decreased temperature on 15/30 (50%) of occasions, cue B (pain exacerbation related cue) was followed by increased temperature on 15/30 (50%) of occasions, and cue C was followed by no change in temperature on 30/30 occasions. The control condition provides additional control in our parametric design, although was initially included to permit a more conventional analysis, (not reported here). The 5 different trial types were presented in random order.

*Behavioural measures*: Subjects performed a reaction time task which consisted of judging whether the visual cue appeared to the left or right of centre on the display monitor, as quickly as possible. The resulting reaction times were taken as a behavioural index of conditioning. Performance on this task was not contingent on the stimuli presented and subjects were told before imaging that their success or failure at quickly judging the position would not affect the amount of pain or relief received. The task was performed with a two-button key-press using the right hand. Heart rate was recorded using a pulse oximeter in conjunction with Spike 2 (CED, Cambridge, UK) software.

A behavioural version of the task was performed that was identical to that performed in the fMRI scanner, only it was performed in a testing room with the subject seated in front of a computer monitor. Following this task, we performed a supplementary cue-preference task, designed to investigate whether the subjects had acquired appetitive and aversive preferences for the cues, as a result of the conditioning procedure. In this task, we presented two cues side-by-side, and asked the subject to make a judgement as to which cue they preferred, by pressing a key-press for left or right. Each cue-pairing was repeated 10 times, and randomised as to which side the cue appeared on. We calculated the preference scores by summing the total number of preference choices made for each cue (as in an all-play-all games table, with a maximum score of 20). Mean scores for each cue were compared across subjects using Wilcoxon sign rank tests.

We did not attempt to formally address the issue of conscious versus non-conscious acquisition of conditioned expectancies. However to gain some insight into the level of explicit expectancy learning, we asked the question ‘Did you recognise any relationship between the pictures and subsequent change in pain level’ at the end of the experiment (for the behavioural version of the task only). Subjects were not told the experiment was a learning / conditioning study beforehand, rather were simply told that it was a study of pain and temperature processing. 10/14 subjects were unable to report any association between cues and outcomes.

*Computational model:* We used a temporal difference model to generate a parametric regressor corresponding to the appetitive prediction error, which was applied to the imaging, as previously described (O'Doherty et al., 2003;Seymour et al., 2004). Here, we used a two time point temporal difference model with a learning rate ($\alpha = 0.3$) determined from behavioural results (see below). In this model, the value $v$ of a particular cue (referred to as a state $s$) is updated according to the learning rule: $v(s) \leftarrow v(s) + \alpha\delta$, where $\delta$ is the prediction error. This is defined as $\delta = r - a + v(s)_{t+1} - v(s)_t$ where $r$ is the return (i.e. the amount of pain) and $a$ is the average amount of reinforcement (tonic pain) that was assumed to be constant. We assigned relief and exacerbations of pain as returns of 1 and -1 respectively (i.e. a linear scale of pain from relief to exacerbation). This is an arbitrary specification, given that is difficult to precisely scale the relative oppositely valenced utilities of relief and exacerbations of pain. Thus, the model treats predictions relating to relief of pain on equal par with unexpected omission of exacerbation of pain; and similarly treats exacerbation related predictions equivalently to unexpected omissions of relief.

*Data acquisition and analysis:* These were taken as measures of cue-reinforcement and correlated with the temporal difference value (i.e. the cue expectancy).

*Reaction time measurements:* Reaction time data were individually (i.e. on a subject by subject basis) fit to a gamma cumulative distribution function (using a maximum likelihood function), to allow analysis across subjects, and correlated with the TD value. This yielded a best fit with a learning rate of 0.3, and a significant correlation with the predicted value (from the model) with both the relief related and exacerbation related trials, independently, and in the same direction. That is, reaction times were shorter for both high reward values and high aversive values. To remove any possible confounding effects of early trials, during which reaction time data habituate substantially, we repeated this procedure after removing the first 10 trials. This yielded a correlation which just failed to reach significance $p=0.056$, across both cue types. We also looked at sensitivity to the TD initial value by setting this to the average value of 0.5, which yielded a non-significant correlation.

*Autonomic:* The heart rate was found to be approximately normally distributed, and was normalised to permit analysis across subjects. We found significant heart rate correlations with both relief and pain cue types (independently, as for the reaction time). For both exacerbation and relief trial types, this yielded a best fit with a learning rate of 0.3. Across both cue types, this remained significant ($p<0.05$, $r=0.19$) after removal of the first 10 trials and with utilisation of different initial TD values. This is a robust correlation, therefore reported in the main text. Consequently we used a learning rate of 0.3 for the TD model used in the fMRI analysis.

*fMRI*. Functional brain images were acquired on a 3T Allegra Siemens scanner. Subjects lay in the scanner with foam head-restraint pads to minimise any movement associated with the painful stimulation. Images were realigned with the first volume, normalised to a standard EPI template, and smoothed using a 6mm FWHM Gaussian kernel. Realignment parameters were inspected visually to identify any potential subjects with excessive head movement, none were found. Images were analysed in an event-related manner using the general linear model, with the onsets of each stimulus represented as a delta function to provide a stimulus function. We employed a parametric design, in which the temporal difference prediction errors modulated the stimulus functions on a stimulus-by-stimulus basis. The statistical basis of this approach has been described previously(Buchel et al., 1998). Regressors were then generated by convolving the stimulus function with a haemodynamic response function (HRF). Effects of no interest included the onsets of visual cues, the pain relief and exacerbations themselves, and realignment parameters from the image pre-processing to provide additional correction for residual subject motion. Linear contrasts of appetitive prediction errors were taken to a group level (random effects) analysis by way of a one-sample t-test, and the aversive prediction error was taken as the inverse. MNI coordinates and statistical z-scores are found in table 1. This analysis determines areas which correlate to univalent appetitive or aversive prediction error, and does not identify areas in which these signals overlap. To explore the possible representation of distinct prediction error signals for the pain relief and exacerbation trials, we generated two independent regressors for the prediction error occurring at each. Then, we took the appetitive relief and aversive exacerbation components of the prediction error to a second level analysis of variance, and exclusively masked the two individual contrasts (ie. looked for areas of overlap of the independent appetitive-relief and aversive-exacerbation prediction errors, both at $p<0.001$). These data are presented in figure 5a-c.

*Anatomical localization and areas of interest*: Group level activations were localized according to the group averaged structural scan. Activations were checked on a subject-by-subject basis using their individual normalised structural scans to ensure correct localization, since some of the reported activations are in small nuclei (e.g. substantia nigra). We report activity in areas in which we had prior hypotheses, based on previous data, though without specification of laterality. These regions have established roles in both aversive and appetitive predictive learning, and included ventral putamen, head of caudate, midbrain (substantia nigra), anterior insula cortex, cerebellum, anterior cingulate cortex, amygdala, lateral orbitfrontal cortex, medial orbitofrontal cortex, dorsal raphe, and ventral tegmental area. We report activations at a threshold of $P<0.001$, with a minimum size of 5 contiguous voxels. We also report brain activations outside our areas of interest that survive whole brain correction for multiple comparisons (see Table1) using family-wise error correction at $p<0.05$.

*Impulse responses*: We performed a supplementary fixed-effects analyses on a trial basis to determine impulse responses, as previously described(Seymour et al., 2004). Note that this analysis refers to the average impulse response across each trial throughout the experiment, and does not embody the time-dependent nature of learning incorporated within the main parametric analysis.

## **4.3 Results**.

*Behavioural measures*. Subjects rated the baseline thermal stimulation as painful, and the decreases and increases in temperature as pleasant or more painful, respectively (see fig 4.2a). In addition, pleasant and pain ratings were significantly greater than equivalent temperature changes on adjacent skin, untreated with capsaicin ($p<0.05$ all pair-wise comparisons)(see methods).

In a behavioural version of the task, outside of the fMRI scanner, we demonstrated conditioning to the relief and exacerbations of pain by following the learning task with a cue-preference task. In this, subjects (n=14) made a forced choice preference judgement of pairs of cues, presented side by side. This revealed a significant preference ordering, with the relief cue preferred to the neutral cue ($p<0.05$, Wilcoxon sign rank test), which was, in turn, preferred to the exacerbation cue ($p<0.01$, Wilcoxon sign rank test)(fig 4.2b). On post-experimental debriefing (see methods), only 4 out of the 14 subjects could report any contingent relationship between the cues and the outcomes.

**Figure 2**

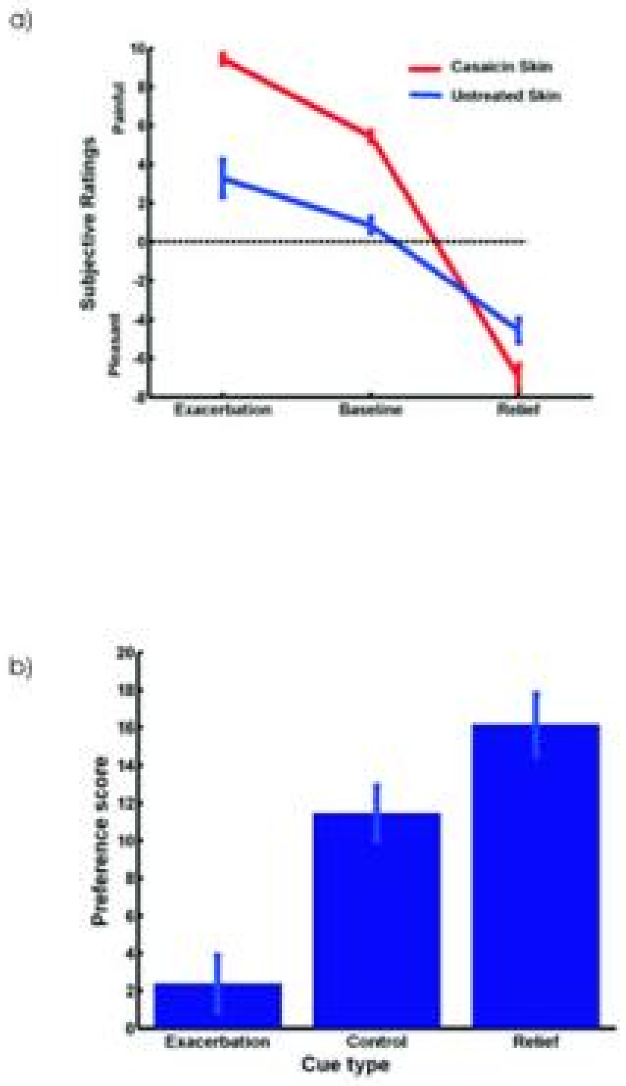
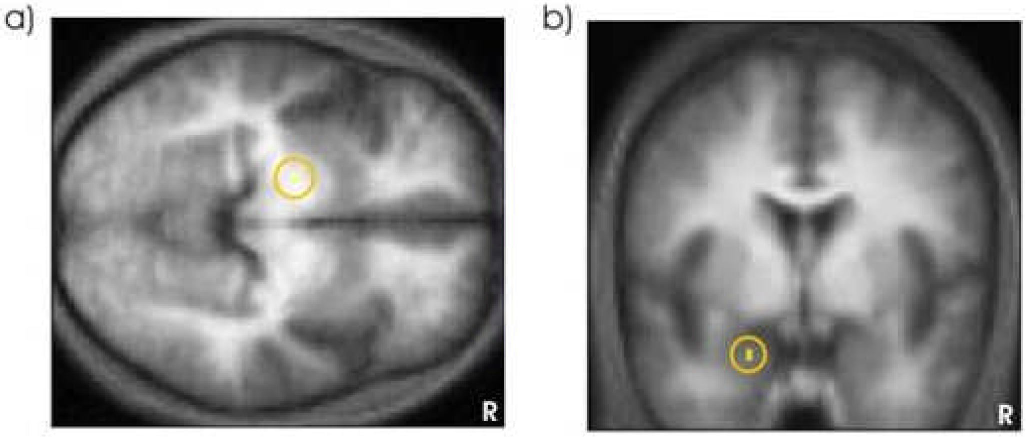


Figure 4.2 **a) Pain ratings.** Pain and pleasantness ratings for the baseline level of thermal stimulation, and the phasic increases and decreases in temperature. Scores are on a 0-10 magnitude rating, with error bars representing the standard error.. The graph shows results for the capsaicin treated skin, and an adjacent area of unaffected skin. **b) Preference scores.** Following the learning experiment, subjects made forced choices between randomised pairs of cues, The scores are out of a maximum of 20 pairings for each cue (with higher scores indicating more preferred).

During the fMRI version of the task, we used physiological measures to assess the acquisition of cue expectations. Heart rate changes induced by the cues correlated with the magnitude of expectations (i.e. cue-specific temporal difference values) both of pain relief (p<0.01) and pain exacerbation (p<0.01), calculated from the model (see methods). This supports the hypothesis that cue expectations are acquired in a manner consistent with the (temporal difference) learning model, albeit in a valence-insensitive manner. That is, we observed increased heart rate with higher valued cues, whether positive or negative, consistent with a learned arousal-like response associated with the expectations.

*Appetitive prediction error*. We used the model to identify a representation of the appetitive prediction error in the brain (see figure 1b, appetitive model). Activity in left amygdala and left midbrain (in a region consistent with the substantia nigra) correlated with this signal (figure 4.3a,b). Time-course analysis illustrates the *average* pattern of response associated with the different trial types in the amygdala, illustrating a strong correspondence with the predictions of the model (figure 4.3c). These data support the hypothesis that relief learning involves a reward-like learning signal.



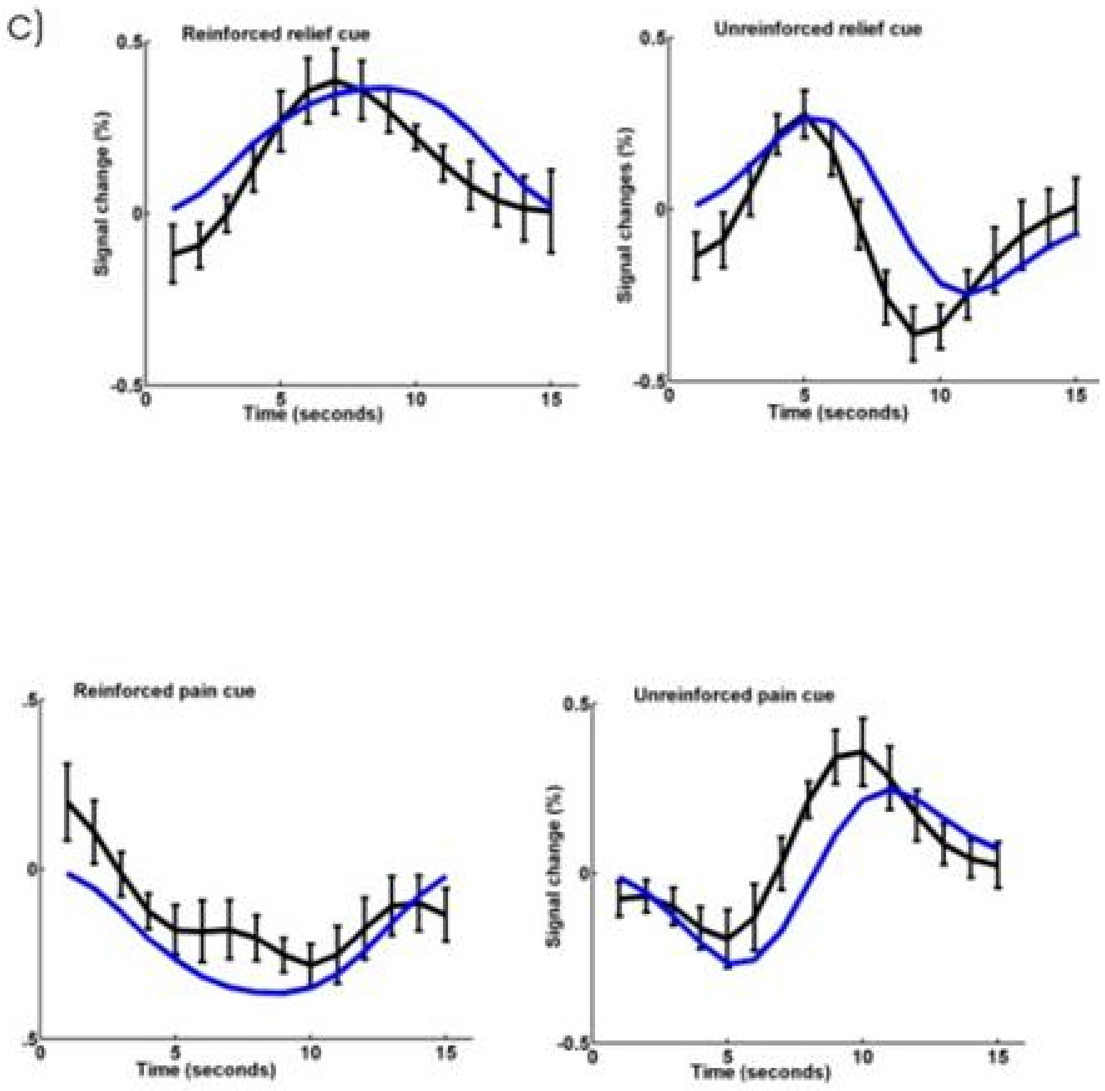
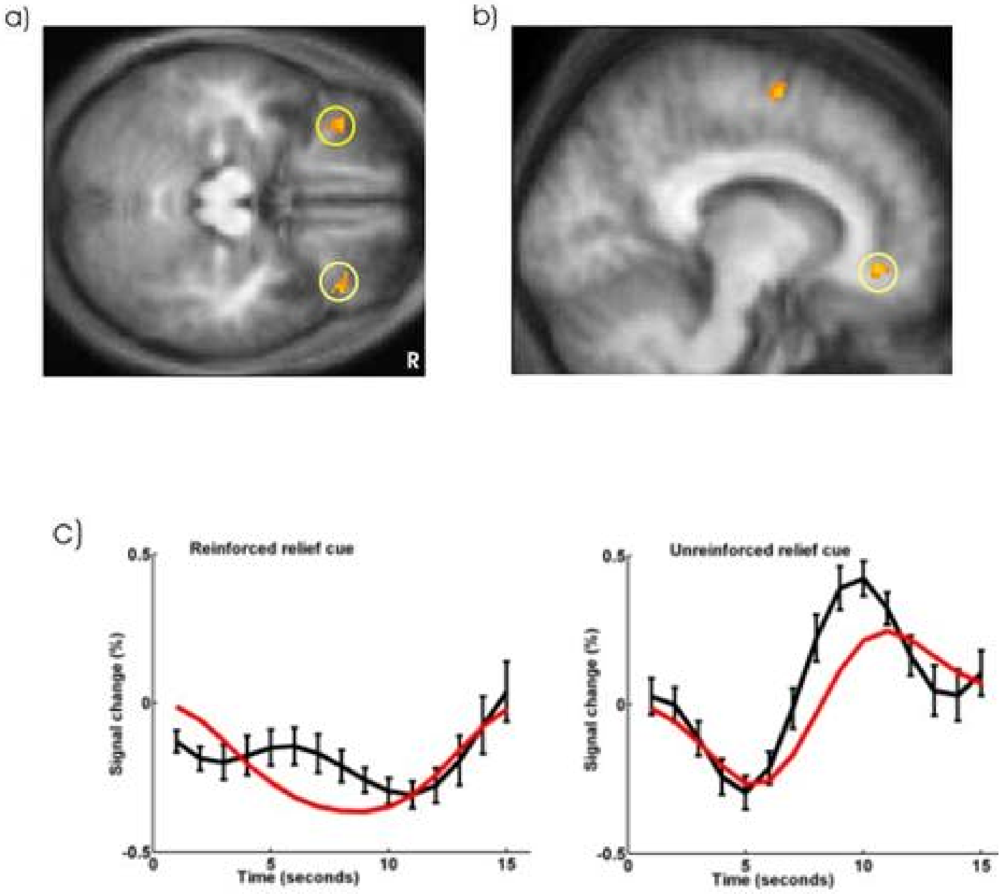


**Figure 4.3 Appetitive temporal difference prediction error.** Statistical parametric maps (p<0.001) showing **a)** left substantia nigra (axial plane) and **b)** left amygdala (coronal plane). **c)** Time course of inferred mean neuronal activity for the four principal trial types in left amygdala. The black line represents the data (error bars represent 1 standard error), and the blue line is the model appetitive temporal difference prediction error (from figure 1b) after convolution with a canonical haemodynamic response function.

*Aversive prediction error*. Recent evidence indicates that the temporal difference model also provides an accurate description of aversive learning, suggesting the existence of a separate learning mechanism that codes for aversive events

(Seymour et al., 2004). We therefore sought to identify whether an aversive representation of the prediction error was expressed, in which exacerbation of pain was treated as positive punishment, and relief as negative punishment (figure 1c, aversive model). Activity in bilateral lateral orbitofrontal cortex and genual anterior cingulate cortex correlated with this signal (fig 4.4a,b). The time-course of this activity, shown in figure 4.4c, illustrates the opposite pattern of response to the appetitive prediction error. These data indicate the existence of an aversive reinforcement signal, distinct from the reward-like signal.



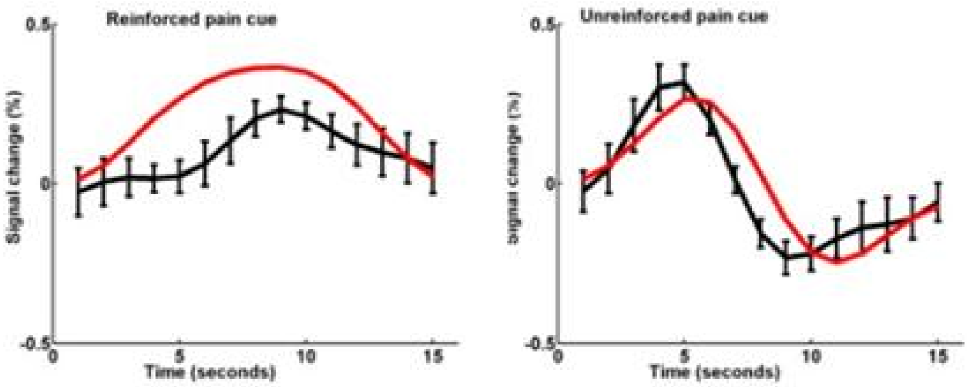
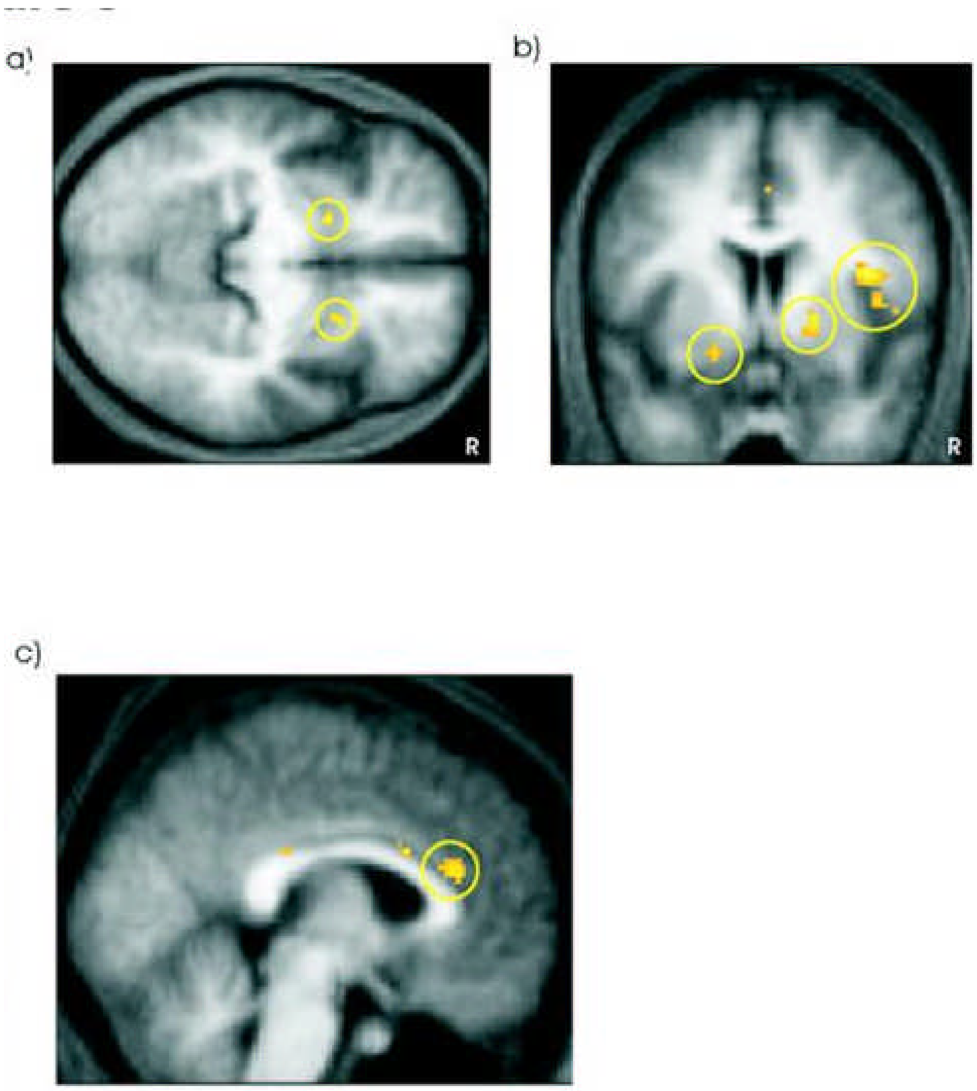


**Figure 4.4 Aversive temporal difference prediction error.** Statistical parametric maps (p<0.001) showing **a)** lateral orbitofrontal cortex (axial plane), and **b)** genual anterior cingulate cortex, highlighted (sagittal plane). **c)** Time course of inferred mean neuronal activity for the four principal trial types in left orbitofrontal cortex. The black line is the data (error bars represent 1 standard error), and the red line is the model aversive temporal difference prediction error (figure 1c) after convolution with a canonical haemodynamic response function.

*Prediction error signal in Ventral Striatum.* Psychological studies of appetitive-aversive interactions predict that opposing, learning related activity should converge in some areas(Dickinson and Dearing MF, 1979). This might occur in areas such as the ventral striatum (and insula cortex), where predictive activity has been observed in both reward and pain learning tasks, albeit in separate studies (Jensen et al., 2003;McClure et al., 2003;O'Doherty et al., 2003;Ploghaus et al., 1999;Setlow et al., 2003;Seymour et al., 2004). This raises a question about how co-expressed aversive and appetitive prediction errors would be represented by the BOLD signal, particularly if they interact. We therefore created a new statistical model that included two regressors, modelling prediction error for relief and exacerbation separately. This model revealed co-expression in the ventral putamen, anterior insula and rostral anterior cingulate cortex (fig 4.5a-c). The responses in these regions showed an appetitive prediction error for the relief related cue, and an aversive prediction error for the exacerbation related cue (fig 4.5d). This pattern of activity is interesting, since it cannot result simply from the linear super-position of appetitive and aversive signals, but implies either an interaction between prediction error and cue-valence, or the expression of a single valence-independent prediction error.



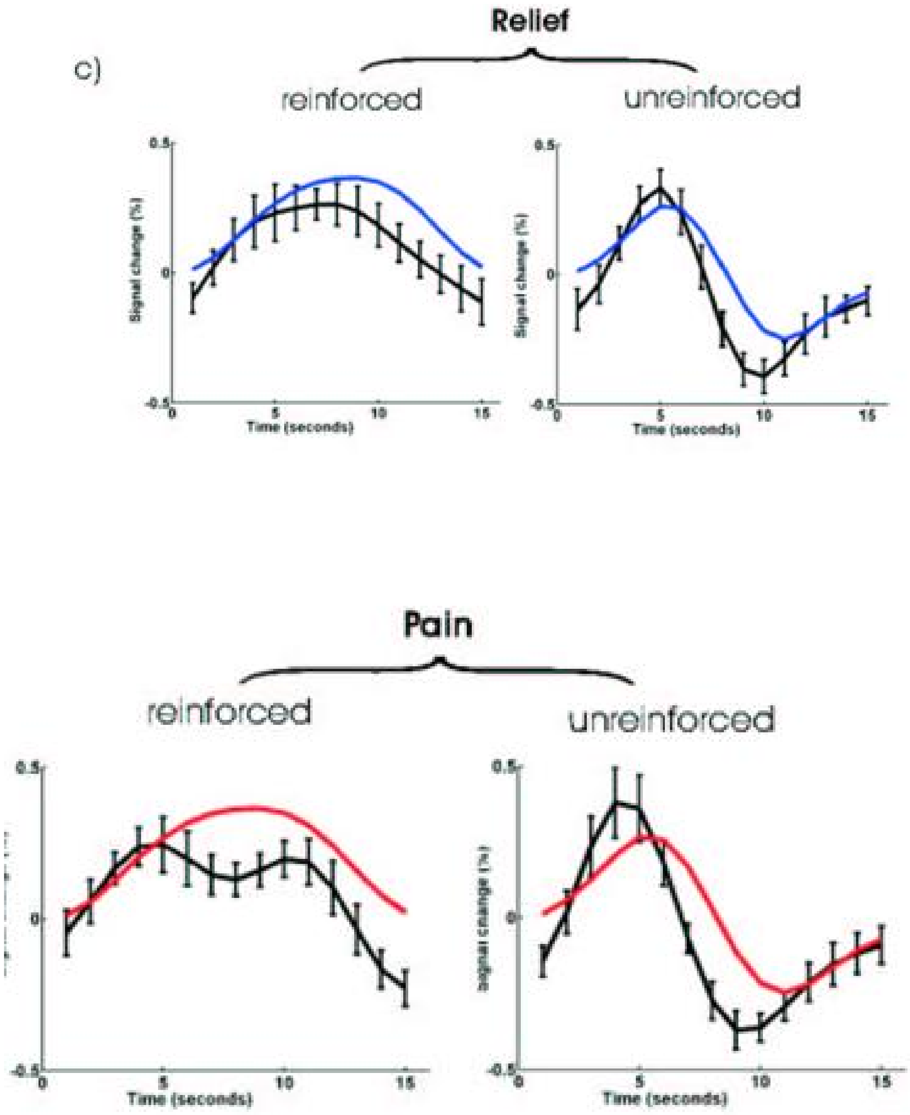


**Figure 5. Appetitive relief-related plus aversive exacerbation-related prediction error**. Statistical parametric maps showing activity that correlates with the appetitive prediction error for the relief cue (p<0.001), masked with the aversive prediction error for the exacerbation cue (p<0.001). **a)** bilateral ventral putamen, **b)** bilateral ventral putamen and right anterior insula **c)** rostral anterior cingulate cortex. **d)** Time course of inferred mean neuronal activity for the four principle trial types in left ventral putamen. The black line represents the data (error bars represent 1 standard error), and the blue and red line is the model appetitive and aversive temporal difference prediction error respectively (from figure 4.1b,c), after convolution with a canonical haemodynamic response function.

## 4.4 Discussion.

Drawing on theoretical considerations provided by reinforcement learning (Daw et al., 2002), we suggest our data provide evidence in support of an opponent model of pain relief. We observed two distinct patterns of neural activity, distinguishable by their expression in separate brain areas, which correlated with the prediction error signals of an opponent temporal difference model. This extends our understanding of human predictive learning beyond the occurrence of simple phasic events arising from a neutral baseline. Thus during tonic pain, aversive and appetitive systems would appear to be simultaneously active to encode appropriate goal-directed predictions across the spectrum of positive and negative outcomes. Our observations provide a formal framework for understanding the homeostatic and motivational processes engaged by pain, and offer a paradigmatic account of motivation during tonic affective states.

The use of the temporal difference algorithm to represent positive and negative deviations of pain intensity from a tonic background level approximates the class of reinforcement learning model termed average-reward models (Daw and Touretzky, 2002;Mahadevan, 1996;Schwartz, 1993). Accordingly, predictions are judged relative to the average level of pain, rather than according to an absolute measure. This comparative treatment of motivationally salient predictions is consistent with both neurobiological and economic accounts of homeostasis, which rely crucially on *change* in affective state (Craig, 2003;Markowitz, 1952).

Implicit in any such model is a representation of the average rate of reinforcement, although the short time window of fMRI precludes investigation of this directly. From an implementational perspective, one argument for opponency relates to consideration of how a long-run average affective state might be represented. Given our demonstration that positive and negative prediction errors are both encoded by one system, and fully mirrored by opposite signals in an opponent system, the requirement for one system to fully represent both the tonic levels of reinforcement (*ie.* by sustained elevated activity) with positive and negative phasic predictions simply superimposed, would appear to be obviated. If this is the case, the tonic level of pain would be free to have a distinct representation, a signal that has been suggested to be conveyed by *tonic* dopamine release (Daw et al., 2002).

Mirror opponency has many similarities to the appetitive-aversive reciprocity characteristic of early psychological ‘opponent process’ theories (Grossberg, 1984;Konorski, 1967;Schull, 1979;Solomon and Corbit, 1974). In their various forms, these theories grew out of a requirement both to explain the adaptive changes that occur during tonic reinforcement (and that follow its termination), and to understand the interactions between appetitive and aversive processes that arise in certain specific learning paradigms such as conditioned inhibition and trans-reinforcer blocking. Interestingly, recent electrophysiological recordings of neuronal activity in mice directly indicate the involvement of opponent processes in (context-related) conditioned inhibition, specifically implicating the ventral striatum and amygdala (Rogan et al., 2005). Thus it seems possible (and fully consistent with a computational account) that, at least in the ventral striatum, a ‘safety-signal’ that predicts the absence of future pain might share the same neural substrate as the relief prediction error seen here. However, we show an appetitive representation in the amygdala, rather than an opponent aversive representation (which we observe instead in lateral orbitfrontal and genual anterior cingulate cortex). This points to the expression of multiple learning-related neural signals in the amygdala, consistent with the complex, integrative role of this structure (and the various nuclei within) in associative learning and pain (Baxter and Murray, 2002;Holland and Gallagher, 2004).

The finding that lateral orbitofrontal cortex demonstrates an aversive prediction error signal is consistent with previous reports of a role for this region in aversive learning (O'Doherty et al., 2001). In particular, this area has been shown to be involved in evaluation of aversive stimuli in the context of different motivational states(Small et al., 2001), as well as in short timescale pain prediction relative to a changing (learned) baseline rate of phasic pain (Glascher and Buchel, 2005b). Taken with the present results, this suggest that learning of aversive value predictions in this region may be mediated by an aversive specific prediction error signal, and particularly in circumstances that require adaptive representations following changing motivational state or context. However, it should also be noted that lateral orbitofrontal cortex may not be exclusively involved in aversive processing, as reward-related responses have also been reported in this region in some circumstances.

In relation to pain, other cortical areas, specifically insula and anterior cingulate cortex, have clear motivational roles in pain and have previously been implicated in the processing of relief-related information (Fields, 2004). For example, recent neuro-imaging studies investigating the expectation and receipt of placebo analgesia implicate these areas in endogenously mediated analgesia (Petrovic et al., 2002;Wager et al., 2004). Our findings provide further support, therefore, that these areas play a key functional role in pain homeostasis (Craig, 2003).

The BOLD signal is thought to correspond to changes (increases or decreases) in synaptic activity, and thus the activity we describe may reflect specific afferent neuromodulatory influences that originate elsewhere (Logothetis et al., 2001;Stefanovic et al., 2004). Substantial evidence indicates that mesolimbic dopamine neurons both encode reward-related prediction error (Dayan and Balleine, 2002;Schultz et al., 1997) and play a key role in analgesia (Altier and Stewart, 1999), suggesting that dopamine could convey an appetitive relief-related prediction error. This draws attention to activity in the ventral striatum, a region that receives strong mesolimbic dopaminergic projections. Comparison with previous data highlights the observation that cues signalling lower-than-predicted pain cause deactivation in this area in the context of a neutral baseline, as opposed to activation in the context of a tonic pain baseline (Jensen et al., 2003;Seymour et al., 2004). This implicates adaptive changes occurring during tonic pain, influencing ventral striatal activity, and consistent with the representation of an appetitive signal for relief related cues. However, taken alone, it is possible that this ventral striatal activity is modulated by a single prediction error signal for both relief and exacerbation cues (Horvitz, 2000;Smith et al., 2005), although recent electrophysiological evidence revealing *suppression* of midbrain dopaminergic neurons to aversive stimuli would seem to require a separate aversive opponent signal (Ungless et al., 2004). Either way, this signal must interact with valence specific information by some additional mechanism, possibly through the involvement of different intrinsic sub-populations of appetitive and aversive neurons within the ventral striatum (Roitman et al., 2005).

That pain relief and reward might share a common neural substrate is also suggested by the fact that many drugs that have rewarding effects have analgesic properties. Aside from dopamine, there are many neurotransmitters with clear combined roles in appetitive and aversive motivation, for example opioid peptides, serotonin, substance P, and glutamate (Fields, 2004;Gadd et al., 2003;Johansen and Fields, 2004). Of particular note are the dorsal raphe serotonergic projections to the ventral striatum, which have been recently proposed to encode the aversive prediction error (Daw et al., 2002).

In addition to a role in Pavlovian motivation, it is also clear that pain and relief-related expectations exert a strong influence on the actual subsequent experience of pain – in that perception (of intensity) is weighted by the prior expectancies acquired through conditioning. Quite how predictive motivational values influence perceptual inferences (such as pain intensity) is not yet clear, although probabilistic perceptual models that incorporate economic cost functions (such as decision theory) may offer insight at a theoretical level(Dayan and Abbott LF, 2001). From an implementational perspective, one putative mechanism exploits an influence of 'higher' brain areas on ascending pain pathways via descending modulatory control centres. A possible target is the 'on-' and 'off-'cells of the periaqueductal grey and rostral ventromedial medulla, which display opponent anticipatory pain related activity under apparent higher control(Fields, 2004). Whatever the mechanisms, these influences are thought to be clinically important both in endogenous pain modulation (including placebo analgesia) and in the pathogenesis of some chronic pain syndromes(Fields, 2004;Petrovic et al., 2002;Price, 1999;Wager et al., 2004), and we suggest that integrated psychological, neurophysiological and computational approaches offer some promise in furthering their understanding.

Recently, Baliki and colleagues (2010) performed an experiment looking at the offset of pain, as well as the onset (the two are de-correlated by varying the duration of a phasic pain stimulus). They did this in chronic back pain subjects, and in healthy controls. What they saw was a clear difference in activity between the two groups at the time of offset: probably the best imaging demonstration to date of differential pain processing in patients and controls. They showed that basic pain activation statistical maps are very similar between groups, but a striking difference in the ventral striatum (a region that seems to include the nucleus accumbens and ventral putamen). At the time of onset of pain, both back pain and control subjects show phasic clear activation of this region. However, at the time of offset of pain, the control patients show a further phasic activation, whereas the back pain patients show a phasic decrease in activity. The authors suggest that the phasic activity at the time of onset may represent a salience or arousal signal associated with the pain in both groups. At the time of offset, they suggest that the control group exhibit an appetitive relief signal, whereas the back pain group exhibit a punishment signal as the patients return to attend to their back pain, manifest negatively in a reward-coding system. As the authors note, the correlation with a derivative of value has parallels with a prediction error. However this raises a couple of awkward problems: salience-based accounts of striatal activity are generally thought of as competing theories of dopaminergic function, rather than in addition to the reward prediction error theories, and so it is difficult to accommodate both accounts within the same pain epoch. Secondly, it is tricky to imagine how a motivational system will consider less pain as punishment, despite the attention-related decrement in back pain during the experimental pain. If this were really the case, then why don't back pain patients seek out phasic pain to distract them from their chronic back pain?

An alternative explanation is that at the time of offset, control subjects adopt a reward-valenced frame, and as such exhibit a dominant appetitive coding of relief, as a 'more reward' prediction error. However the back pain patients have a persistently aversive baseline, and so exhibit a dominant aversive representation of relief, as 'less punishment', coded as an aversive prediction error. What is needed to resolve these different interpretations is some way of pharmacologically or anatomically dissociating appetitive and aversive pathways within the ventral striatum.

**Table 4.1.** MNI coordinates and statistical z-scores for the appetitive, aversive and joint co-expressed appetitive-aversive temporal difference prediction error.

| Region | Laterality | X | Y | Z | z-score |
|---|---|---|---|---|---|
| ***Appetitive prediction error*** | | | | | |
| Midbrain (Substantia nigra) | L | -18 | -12 | -8 | 3.99 |
| Amygdala | L | -20 | 2 | -26 | 3.33 |
| ***Aversive prediction error*** | | | | | |
| Lateral orbitfrontal cortex | R | 40 | 34 | -20 | 3.72 |
| | L | -34 | 34 | -20 | 3.71 |
| Genual anterior cingulate cortex | R | 10 | 42 | -6 | 4.24 |
| Motor cortex | R | 14 | 0 | 60 | 5.35[¶] |
| ***Combined appetitive-aversive prediction error*** | | | | | |
| Ventral putamen | R | 18 | 8 | 0 | 4.08 |
| | | 22 | 10 | -10 | 3.32 |
| | L | -18 | 8 | -12 | 3.62 |
| Anterior insula | R | 30 | 22 | 6 | 3.87 |
| | | 36 | 2 | 16 | 4.78 |
| | L | -34 | 12 | 12 | 4.55 |
| Rostral anterior cingulate cortex | R | 2 | 34 | 20 | 3.61 |

[¶] Significant following whole brain correction

# Chapter 5. Differential striatal activity underlies appetitive and aversive learning for monetary gains and losses (experiment 3).

## 5.1 Introduction.

A wealth of human and animal studies implicates ventral and dorsal regions of the striatum in aspects of the learned control of behaviour in the face of rewards and punishments. In experiments involving primary rewards and punishments, the BOLD signal in the human striatum measured using fMRI covaries closely with key learning signals employed by abstract learning models (Haruno et al., 2004;O'Doherty et al., 2003;Seymour et al., 2004;Tanaka et al., 2004;Tanaka et al., 2006;Yacubian et al., 2006). These algorithms originate in sound psychological learning accounts, and are known to acquire normative predictions and affectively optimal behaviours (Barto, 1995;Sutton RS and Barto AG, 1990;Sutton and Barto, 1981).

However two, related, sets of findings, regarding the orientation of this signal and the relationship between rewards and punishments, remain difficult to accommodate fully under this interpretation. First, the BOLD signal seen in the striatum typically takes the form of a signed prediction error, with baseline activity when outcomes match their predictions, and above- and below-baseline excursions when outcomes are more or less than expected, respectively. Of course, rewards and punishments have opposite valences, with a *negative* punishment (e.g., one expected but omitted) bearing a close computational and psychological relationship with a *positive* reward. However, in experiments that involve cues that predict exclusively rewards (which can be presented or omitted), or exclusively primary punishments (which can also be presented or omitted), the BOLD signals are apparently oppositely oriented, with positive BOLD excursions accompanying both positive reward and positive punishment, and below-baseline excursions accompanying both negative (or omitted) reward and punishment (Becerra et al., 2001;Breiter et al., 2001;Delgado et al., 2000;Elliott et al., 2003;Jensen et al., 2003;Knutson et al., 2000;Nieuwenhuis et al., 2005;O'Doherty et al., 2003;Pagnoni et al., 2002;Seymour et al., 2004;Seymour et al., 2005;Tanaka et al., 2004;Yacubian et al., 2006;Zink et al., 2003).

Second, in the above experiments that involve financial costs (in contrast to those involving primary punishments such as physical pain), the striatal BOLD signal is typically observed to be oriented as in rewarding tasks, with monetary gains associated with positive BOLD activations, and losses with *sub*-baseline signals. Indeed, there are few reports of *any* brain areas showing a positive BOLD response to financial loss at all, and although this is not exclusively the case (for instance in amygdala for instance (Yacubian et al., 2006), and insula cortex (Knutson et al., 2007a), it has been suggested that monetary losses and gains might be fully processed by a unitary (appetitive) system, centred on the striatum (Tom et al., 2007).

Potential explanations for these puzzles include the possibility that the striatal BOLD signal reflects the release of different neuromodulators (Daw et al., 2002;Doya, 2002)(one reporting prediction errors of each valence), or the possibility that that neighbouring regions of the striatu m report on the different valences (Reynolds and Berridge, 2001;Reynolds and Berridge, 2002). Indeed, there are sound psychological and neurophysiological reasons to think that separate, opponent systems are responsible for the two valences (Dickinson and Dearing MF, 1979;Gray, 1991;Konorski, 1967). But on this interpretation it remains unclear why different circumstances implicate each signal – for instance, why pain is apparently reported by a punishment-oriented prediction error, but monetary losses are not. We designed a Pavlovian conditioning experiment, involving mixed gain and loss outcomes, to address these underlying issues.

The key requirements for the task were to integrate monetary predictions about gains and losses, and to avoid framing the problem entirely in terms of one valence. One strategy for mitigating the latter, at the potential expense of low experimental power and only subtle outcomes, is to make the task involve predictions alone, with no requirement for action, and so avoiding subjects having expectations that they will be able to win. Thus, we used functional magnetic resonance imaging (fMRI) to examine striatal representations of financial loss in tasks which involve mixed gains and losses, using a probabilistic first-order Pavlovian learning task with monetary outcomes. Importantly, the design included both mixed and non-mixed valence outcome probabilities, allowing us to look specifically at the influence on outcome representations (specifically, the prediction error) of the context provided by the non-experienced outcome (figure 5.1).

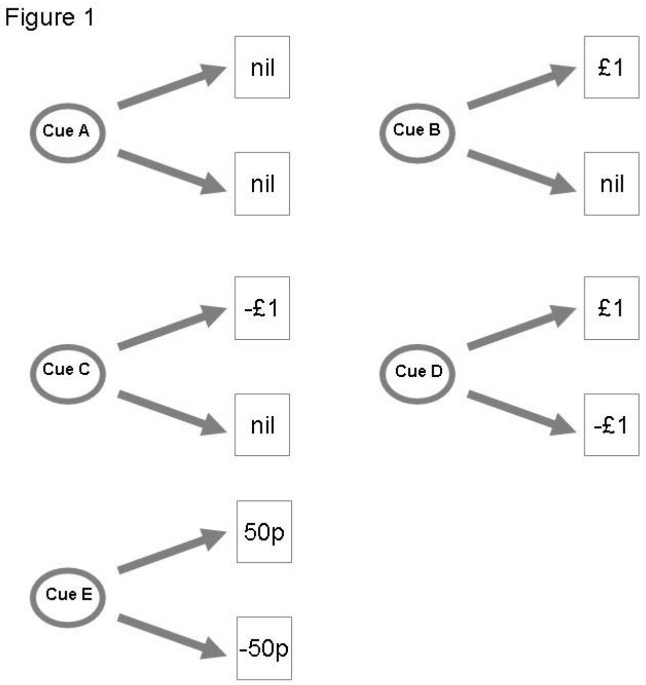


**Figure 5.1. Experimental design.** Visual cues were presented for 3.5 seconds, and followed immediately with the outcome, displayed for 1.5 seconds, depicting the outcome amount. For the analysis, events were marked at the time of the outcome, and linear contrasts performed between the different outcome types.

## 5.2. Methods

*Subjects:* Twenty four (11 female) subjects, age range 19-35, participated in the study. All were free of neurological or psychiatric disease, and fully consented to participate. The study was approved by the Joint National Hospital for Neurology and Neurosurgery (UCLH NHS trust) and Institute of Neurology (UCL) Ethics Committee. Subjects were remunerated by amounts corresponding to their actual winnings during the task (mean zero), added to a fixed pre-stated amount for time and inconvenience (£20).

*Stimuli and Task:* We performed a probabilistic first order Pavlovian delay conditioning task, with visual cues predictively paired with monetary outcomes, as demonstrated in figure 1. Visual cues were presented on a computer monitor projected onto a screen, visible via an angled mirror on top of the fMRI headcoil. The visual stimuli were presented for 3.5 seconds, and on termination were followed immediately by a 1.5 sec duration image of their outcome, either an empty circle (no outcome), a 50pence or £1.00 coin, below which was written in bold letters the amount, and whether they had won or lost (for example 'WIN £1.00'). The 5 cues predicted the following outcomes:

| Cue | Outcome | Probability |
|---|---|---|
| Neutral | £0 | 1 |
| Univalent reward | £0<br>£1 | 0.5<br>0.5 |
| Univalent loss | £0<br>£-1 | 0.5<br>0.5 |
| Bivalent cue (£1) | £1<br>£-1 | 0.5<br>0.5 |
| Bivalent cue (50p) | £0.50<br>£-0.50 | 0.5<br>0.5 |

The visual stimuli were abstract coloured images, approximately 6cm in diameter viewed on the projector screen from a distance of approximately 50cm. They were fully balanced and randomised across subjects, and matched for luminance. We presented 200 trials over 2 sessions, with each trial being presented with a jittered interval of 2-6 seconds.

*Preference task:* Following the conditioning task, we assessed the acquisition of Pavlovian cue values using a preference task, involving forced choices between pairs of cues. Each cue was presented alongside (horizontally adjacent) each other cue, and subjects (still inside the fMRI scanner) made an arbitrary preference judgement between them, using a response keypad (no outcomes were delivered). Each possible combination was presented 5 times (making 50 trials), in random order, and with the position of each cue (on the left or right side of the screen) also randomised. The total number of preference choices for each cue was summed (in a similar manner to a league table) and non-parametric comparisons assessed statistically.

*Pupillometry:* Pupil diameter was measured online during fMRI scanning by an infrared eye tracker (Applied Sciences Laboratories, Waltham MA, Model 504) recording at 60 Hz. Pupil recordings were analysed on an event-related trial basis, and used to find evidence of basic conditioning between the reward, aversive and neutral cue. We used the peak light reflex following presentation of the cue, which is a standard measure of autonomic arousal (Bitsios et al., 2004), and we performed analyses using a repeated measures ANOVA and post hoc t-tests. Technical problems led to the data not being collected for 4/24 subjects.

*fMRI*: Subjects learned the task *de novo* in a functional magnetic resonance imaging (fMRI) scanner to allow us to record regionally specific neural responses. Functional brain images were acquired on a 1.5T Sonata Siemens scanner. Subjects lay in the scanner with foam head-restraint pads to minimise any movement. Images were realigned with the first volume, normalised to a standard EPI template, and smoothed using a 6mm FWHM Gaussian kernel. Realignment parameters were inspected visually to identify any potential subjects with excessive head movement, none was found. Images were analysed in an event-related manner using the general linear model, with the onsets of each outcome represented as a delta function to provide a stimulus function. Regressors of interest (10 in total) were then generated by convolving the stimulus function with a haemodynamic response function (HRF). Effects of no interest included the onsets of visual cues and realignment parameters from the image pre-processing to provide additional correction for residual subject motion.

Linear contrasts of the outcomes SPMs were taken to a group level (random effects) analysis by way of a one-sample t-test. MNI coordinates and statistical z-scores are reported in figure legends.

Group level activations were localized according to the group averaged structural scan. Activations were checked on a subject-by-subject basis using individual normalised structural scans, acquired after the functional test scanning phase, to ensure correct localization. We report activity in areas in which we had prior hypotheses, based on previous data, though without specification of laterality. These regions have established roles in both aversive and appetitive predictive learning, and included putamen, caudate, nucleus accumbens, midbrain (substantia nigra), amygdala, anterior insula cortex, and orbitfrontal cortex. We report activations at a threshold of P<0.001, which survive false discovery rate (FDR) correction at p<0.05 for multiple comparisons using a 8mm sphere around coordinates based on previous studies. Note that in the figures (3 and 4) we use a threshold of p < 0.005 (with a 5 voxel extent threshold) for display purposes. No other activation was found outside our areas of interest that survived whole brain correction for multiple comparisons using FDR correction at $P$<0.05. Details and statistics of all significant activations appear in the figure legends of the appropriate contrasts.

We performed two central analyses. One involved trial-based contrasts for positive reward and loss prediction errors:

i) Positive reward prediction error: bivalent £1.00 win outcome minus univalent £1.00 win outcome.

ii) Positive loss prediction error: bivalent £1.00 loss minus univalent £1.00 loss outcome.

In the second analysis, we used a simple reinforcement learning model to generate a signal corresponding to the outcome prediction error, which, as in previous studies, was applied as a regressor to the imaging data(O'Doherty et al., 2003). Here, we used a temporal difference model with a learning rate $\alpha = 0.3$ based on our previous data from Pavlovian learning(Seymour et al., 2005)(although note that the results presented below are robust to changes in learning in realistic ranges (0.3 -0.7)). In this model, the value $v$ of a particular cue (referred to as a state $s$) is updated according to the learning rule: $v(s) \leftarrow v(s) + \alpha\delta$, where $\delta$ is the prediction error. This is defined as $\delta = r_t - v(s)_t$ where $r$ is the return (i.e. the amount of money). We employed a parametric design, in which the temporal difference prediction error modulated the stimulus functions on a stimulus-by-stimulus basis. The statistical basis of this approach has been described previously(Buchel et al., 1998;O'Doherty et al., 2003). Regressors corresponding to the outcome prediction errors were then generated by convolving the stimulus function with a haemodynamic response function (HRF).

Finally, we considered two further trial-based contrasts. One sought the representation of the negative prediction errors:

iii) Negative reward prediction error: univalent £1.00 win outcome minus bivalent £1.00 win outcome.

iv) Negative loss prediction error: univalent £1.00 loss minus bivalent £1.00 loss outcome.

These contrasts afforded no significant difference at our thresholds.

The second contrast considered residual activity in striatum, when equal prediction errors are subtracted:

v) Zero net prediction error: bivalent £0.50 win outcome minus univalent £1.00 win outcome.

vi) Zero net prediction error: bivalent £0.50 loss outcome minus univalent £1.00 loss outcome.

As expected from standard models, none of these contrasts yielded a significant difference.

To address the possibility that cue-related responses might confound identification of prediction error related responses, we repeated all analyses (both trial-based and model-based), with the inclusion of a single cue related regressor. Inspection of the regressor covariance matrix relating to parameter estimability following convolution of the design matrix with the HRF suggested that the models were not over-specified. Indeed, for the model-based analysis, there was no correlation between the cue regressor and the prediction error. In keeping with this, the SPMs for both trial-based and model-based analyses showed minimal changes in results. Second, we repeated the trial-based analysis with full specification of the identity of the cue, ie, with 5 separate cue regressors. As above, this did not alter the results to any substantial degree. Third, we orthogonalised the outcome regressors with respect to the cue regressors, and again, the results changed only minimally (in either direction). No significant correlations were found with the cue-related regressors.

## 3. Results.

*Behavioural results:* The post-conditioning preference task demonstrated significant preference for the cue associated with univalent reward cue over the neutral cue, in turn preferred to the univalent loss cue. Preference scores for the bivalent cues were slightly above those of the neutral cue, for which the expected value is equivalent (see figure 5.2; see figure legend for statistics). Pupil diameter, which is an autonomic measure of arousal, also provided evidence of basic conditioning to the rewarding and aversive cues, compared to neutral cue (see figure 5.2; see figure legend for statistics).

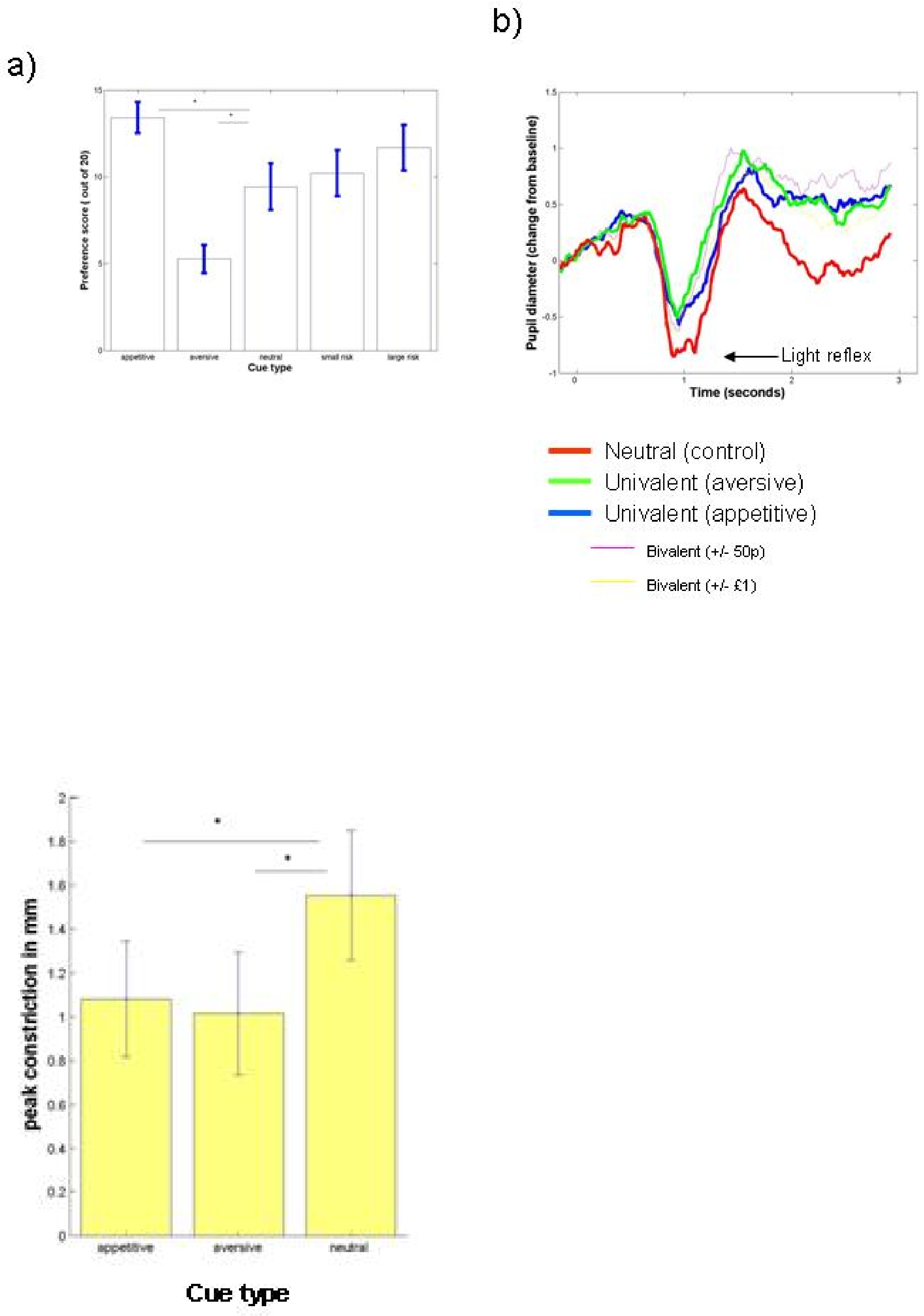


**Figure 5.2 . Behavioural results**. a) Preference scores: One-way repeated measures anova F(4,92)=5.572 p=0.0005; post-hoc two-tailed t test yielded significant differences between univalent reward and neutral, and univalent loss and neutral (p<0.05). b) Mean pupillometry, average across all trials across learning, in a trial specific manner. We looked for a basic effect of conditioning between the rewarding, aversive and neutral cue, which is a standard measure of conditioning. Repeated measures ANOVA revealed a significant effect of trial type F(2,19)=3.342, p<0.05, and post-hoc t-tests showed a significant effect

(increased amplitude of light reflex) for both rewarding and aversive cues when compared to the neutral cue (p<0.05).

*fMRI results*: The experimental design allowed comparison of neural responses to winning money in two conditions: one in which the alternative was winning nothing, and one in which the alternative was losing. Similarly, it allows comparison of neural activity corresponding to losing money when the alternative was nothing, or winning.  Thus, the key BOLD contrasts were between the univalent and bivalent outcomes, for both gain and loss outcomes, since these reveal appetitive and aversive (respectively) prediction errors specifically relating to the outcomes associated with mixed-valence predictions.

In the appetitive case [bivalent cue followed by £1 reward – univalent reward cue followed by £1 reward] this corresponds to a positive relative reward prediction error of 50 pence, and was associated with activation in ventral striatum (see figure 5.3a). In the aversive case [bivalent cue followed by £1 loss – univalent loss cue followed by £1 loss], this corresponds to a positive aversive prediction error of -50 pence, and was also associated with activation in ventral striatum (figure 5.3b). The peak of the aversive prediction error was slightly posterior to the appetitive prediction error, as shown in the sagittal section displayed in fig 5.3c.

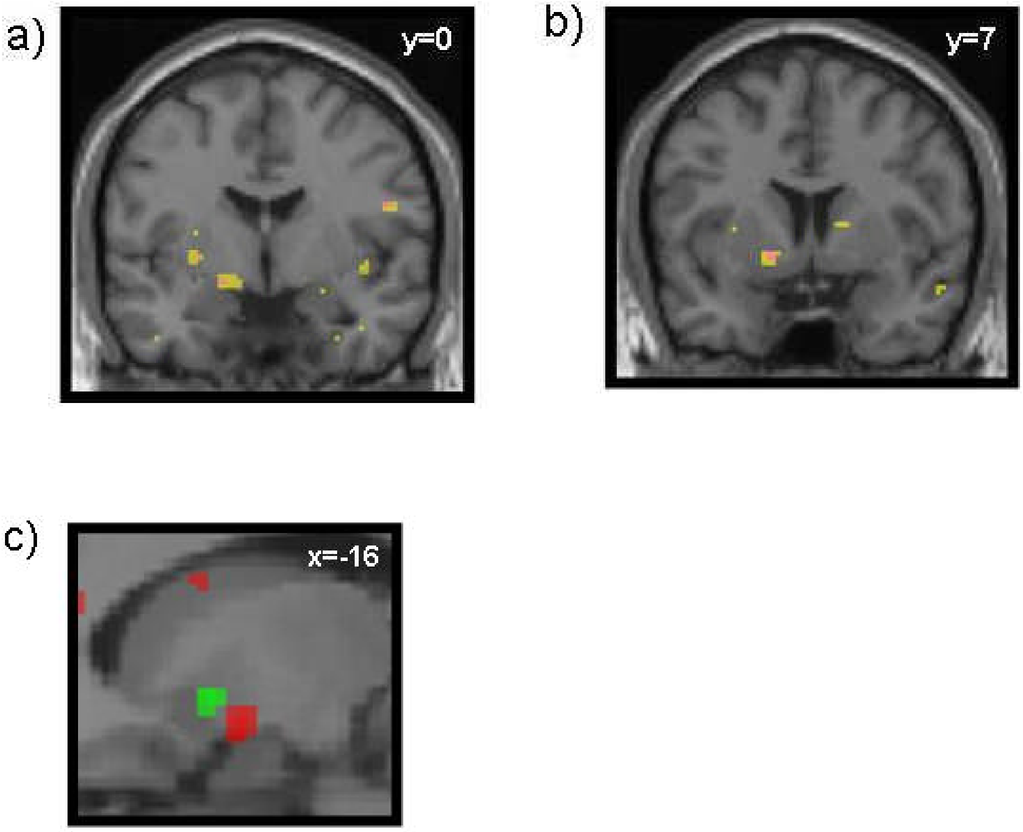


Figure 5.3. **fMRI simple bivalent – univalent contrasts**. a) aversive prediction error: right ventral striatum -16 0 -10, z = 3.74, 46 voxels at p < 0.005. This contrast also revealed a peak in right anterior insula (not shown, 30 18 -12, z = 3.60). Yellow corresponds to p<0.005, magenta corresponds to p<0.001. b) reward prediction error: right ventral striatum -16 6 -6, z = 3.38, 28 voxels at p < 0.005. Yellow corresponds to p<0.005, magenta corresponds to p<0.001. c) sagittal view showing the two peaks: reward (green) and aversive (red).

However, the magnitude of these peaks was such that this analysis could not reliably differentiate the location of appetitive and aversive prediction errors, with the activity in each peak being only insignificantly greater than activity associated with the contrast that defined the other peak. Further, the trial-based contrasts (iii and iv) testing for negative prediction errors of either valence showed no significant effects. This could reflect an asymmetry reported at the spiking level for dopamine neurons (Bayer and Glimcher, 2005;Fiorillo et al., 2003;Morris et al., 2006;Niv et al., 2005), where positive errors are coded more strongly than negative ones. It may additionally be due to the relatively crude trial-based measures.

Therefore, we considered a more sensitive analysis based on a temporal difference learning model. This model is known to offer a good account of the neurophysiological responses of dopamine cells associated with Pavlovian learning about rewards in monkeys (Montague et al., 1996;Schultz et al., 1997), and has been successfully used in human fMRI to probe prediction error components of the BOLD signal from the striatal targets of these cells (Haruno et al., 2004;O'Doherty et al., 2003;Seymour et al., 2004;Tanaka et al., 2004;Tanaka et al., 2006). We applied the model as in previous studies, and used the prediction error occurring at the time of the outcomes generated by this model as a parametric regressor in the fMRI data analysis. This model incorporates both positive and negative prediction errors, and thus identifies valence specific responses. Aversive prediction errors should be negatively correlated with this signal; appetitive prediction errors should be positively correlated with it. Therefore, unlike the trial-based contrasts, this analysis should identify areas that are specific to either valence.

In other words, this analysis identifies subject-specific, trial-specific activity that correlates with the prediction errors fitted by the temporal difference learning model. This analysis was applied solely to the bivalent cues (since it is during these trials that we expected to find opponent prediction error representations).

Activity associated with an aversive temporal difference outcome prediction error was observed posteriorly in the mid putamen (fig 5.4a). Activity associated with an appetitive temporal difference outcome prediction error was observed in more anterior ventral striatum, in close proximity to the nucleus accumbens (see figure 5.4b). These activations are presented in sagittal sections (fig 5.4c; green and red respectively), to permit comparison with the simple prediction error contrasts shown in fig 5.3c.

Given a recent report of identification of an aversive prediction error in amygdala (Yacubian et al., 2006), we looked at a reduced threshold (uncorrected p<0.01) specifically in region. However, no correlated activity was identified.

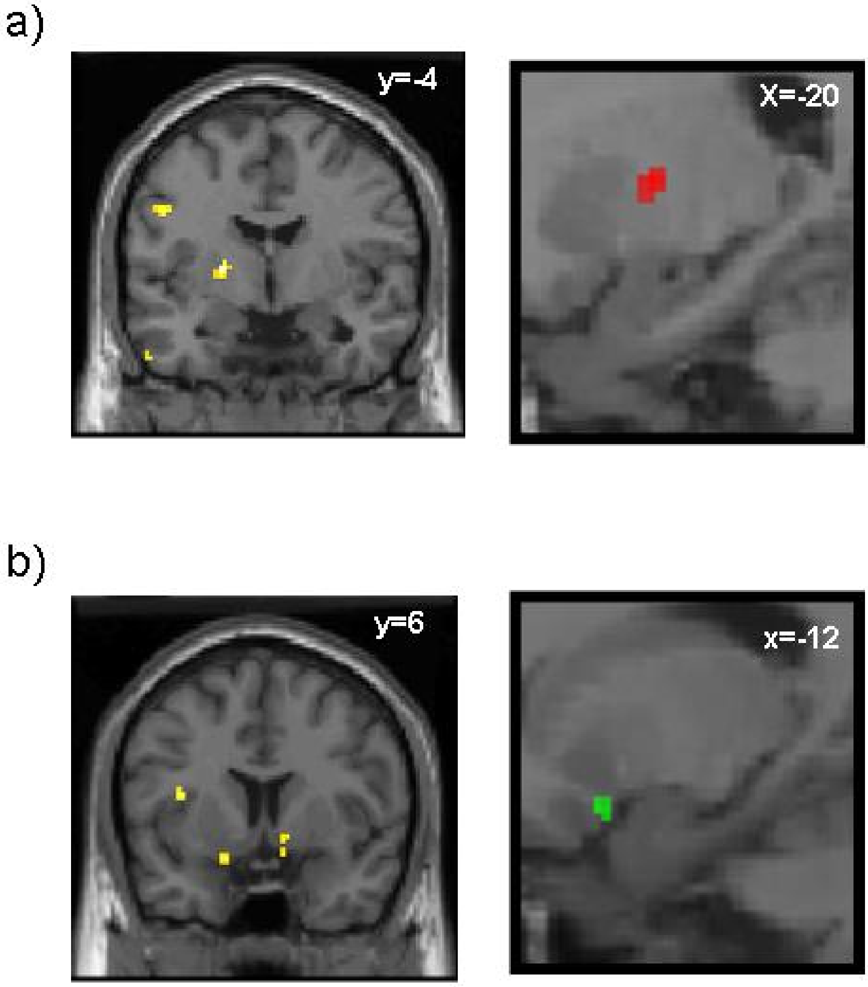


**Figure 5.4. fMRI TD model**: a) aversive TD error: right mid striatum (MNI coordinates: -20 -4 6; z = 3.89, p<0.005, 21 voxels). Yellow corresponds to p<0.005, magenta corresponds to p<0.001. Shown also in sagittal section, in red (right). b) appetitive TD error: right ventral striatum (nucleus accumbens): MNI coordinates: 10 6 -1; z = 3.13, shown at p < 0.005, 15 voxels); left ventral striatum (nucleus accumbens): MNI coordinates: -12 6 -18 z = 3.62, 14 voxels). Yellow corresponds to p<0.005, magenta corresponds to p<0.001. Shown also in sagittal section, in green (right).

## 5.4. Discussion.

Our results suggest a partial resolution to the puzzles outlined in the introduction. The data suggest that aversive and appetitive prediction errors may be represented in a similar manner, albeit somewhat spatially resolvable along an axis of the striatum. The appetitive prediction error appears to direct the BOLD signal in more anterior and more ventral regions than the aversive prediction error. Furthermore, it appears that the prevalence of each sort of coding may depend on the affective context.

Although one should be cautious regarding the topographic spatial resolution of fMRI, the anterior-posterior gradient resembles that seen in stimulation studies of the ventral striatum in rats, in which micro-injecting a GABA agonist or a glutamate antagonist into more anterior regions produces appetitive responses (feeding), and into more posterior regions, produces aversive responses (paw treading, burying) (Reynolds and Berridge, 2001;Reynolds and Berridge, 2002;Reynolds and Berridge, 2003). These studies are characteristic of a growing body of evidence pointing to role of the ventral striatum in aversive motivation, and with distinct neuronal responses associated with appetitive and aversive events (Levita et al, 2002;Horvitz, 2000;Ikemoto and Panksepp, 1999;Jensen et al., 2003;Roitman et al., 2005;Schoenbaum and Setlow, 2003;Setlow et al., 2003;Seymour et al., 2004;Seymour et al., 2005;Wilson and Bowman, 2005).

Aversive learning is well recognised to involve the amygdala. Interestingly, a recent gambling study involving mixed gains and losses of money, at differing amounts and probabilities, identified loss prediction errors in the amygdala, but only gain related prediction error in the striatum (Yacubian et al., 2006). Although it is difficult to place too much emphasis on the respective negative findings for this and our studies, it is noteworthy that these two areas are richly interconnected, both directly and indirectly (Russchen et al., 1985).

The anatomical separation within the striatum could well be *accompanied* by a separation in terms of the relevant neuromodulators (Daw et al., 2002;Doya, 2002). A substantial body of data points to the role of dopamine in striatal reward related activity (Everitt et al., 1999;Montague et al., 2004;Wise, 2004), specifically relating to the representation of prediction errors that guide learning in Pavlovian and instrumental learning tasks (Montague et al., 1996;Schultz et al., 1997). Furthermore, dopamine has been observed to modulate striatal reward prediction errors in human monetary gambling tasks selectively (Pessiglione et al., 2006). If dopamine is involved in the appetitive prediction error observed here, this raises the question as to nature of the aversive prediction error signal, given previous observations and current controversies concerning dopaminergic involvement in aversive behaviours (Horvitz, 2000;Ikemoto and Panksepp, 1999;Ungless, 2004). One possibility is that serotonin released from the dorsal raphe nucleus plays this role (Daw et al., 2002). Consistent with this hypothesis, there is evidence of serotonin-dopamine gradient along a caudal-rostral axis in the striatum (Brown and Molliver, 2000;Heidbreder et al., 1999). However, since our study was not pharmacological, we cannot rule out the possibility that, instead of there being a separate, non-dopaminergic opponent, dopamine provides a valence-independent signal that interacts with valence-specific activity intrinsically coded in striatum (Seymour et al., 2005).

From the perspective of studies into financial decision making and prediction, it is noteworthy that we see striatal BOLD signals above baseline associated with prediction errors for financial losses, whereas most previous imaging studies involving positive and negative financial returns show only decreases below baseline for unexpected losses. This result is important since it makes the findings for financial losses consonant with those for primary aversive outcomes such as pain. It also reinforces caution in the interpretation of striatal activity in human decision making tasks, which as noted in the past (Poldrack, 2006) are sometimes prone to the reverse inference that striatal activity implies the operation of reward mechanisms.

One possible reason for the difference between our and previous results is that in experimental monetary decision making tasks, subjects make choices under the reasonable expectation (perhaps based on implicit knowledge of the mores of ethical committees) of a net financial gain. This establishes an appetitive context or frame within which all outcomes are judged. By contrast, most decisions in day-to-day life involve risks that span positive and negative outcomes; we hoped that mixed-outcome prediction, with no opportunity for choice, would avoid such a frame. Empirical work in finance and economics has suggested that such mixed-outcome decisions fit rather awkwardly within the descriptive framework usually applied to decisions that involve pure gains and losses. Constructs such as Prospect theory suggest a strong dependence of decision-making on valence-context (positive or negative) in which options are judged (Levy and Levy, 2002). The absence of a positive orientation for loss prediction errors in previous studies may thus have arisen from such positive frames. Our results hint that more naturalistic human studies that involve genuine risk of financial loss may be critical to gain further insights into the role of the striatum and other structures into the judgement and integration of gains and losses.

# Chapter 6: Instrumental learning for rewards and punishments, and the role of serotonin (experiment 4).

## 6.1 Introduction

Despite a wealth of data implicating serotonin in motivated behaviour and decision-making, it has been remarkably difficult to identify the precise computational functions that it mediates. Existing theories propose roles in aversive learning and reward-punishment opponency (both phasic and tonic), behavioural flexibility, time discounting, and behavioural inhibition(Cools et al., 2008;Daw et al., 2002;Dayan and Huys, 2008;Doya, 2002;Robbins and Crockett, 2009). Indeed one of the notable and consistent observations from human and animal studies of decision-making is the persistence in choosing options that offer dwindling returns or even intermittent punishment that occurs when central levels of serotonin are depleted(Walker et al., 2009). This seems likely to reflect a core process by which serotonin controls choice, but it could in principle relate to any number of distinct mechanisms also associated with serotonergic function, such as impaired representation or impaired learning about either obtained or omitted rewards, or punishments, or some other aspect of behavioural flexibility. To date, reinforcement learning theory has proved remarkably useful in pulling apart different components of decision-making, offering an accurate account of both behavioural and neurophysiological data (Daw and Doya, 2006). The paradigmatic example is serotonin's companion monoamine neuromodulator, dopamine, which plays an increasingly well-understood role in reward learning (Montague et al., 2004;Schultz et al., 1997). Indeed, it has been proposed that serotonin serves in some fashion to oppose dopaminergic signaling. However, even that mechanistic possibility suggests at least two potential computational functions – either signalling punishments, or signalling an average reward level against which outcomes are weighed – and these have been hard to distinguish, at least in part for the methodological reason that existing tasks have not been able to selectively probe distinct computational aspects of reward and punishment learning.

Here, we used simultaneous instrumental reward and avoidance learning in a four-armed bandit paradigm (figure 1), and probed the contribution of serotonergic mechanisms using acute dietary tryptophan depletion. On each trial, subjects (n=30) selected one of four possible actions, each of which was associated with some chance of reward (20 pence) and also some chance of punishment (a painful electric shock). Importantly, on each trial, each outcome was delivered, or not, according to an independent random choice – like two coin flips – allowing us to unambiguously determine their effects on choice behaviour and neural activity. The probabilities of reward and punishment were independent from one another and also independent between machines, and evolved slowly over time between zero and 0.5 according to separate random diffusions. This required subjects constantly to relearn the values of each bandit, and balance information acquisition (exploration) with reward acquisition and punishment avoidance (exploitation).

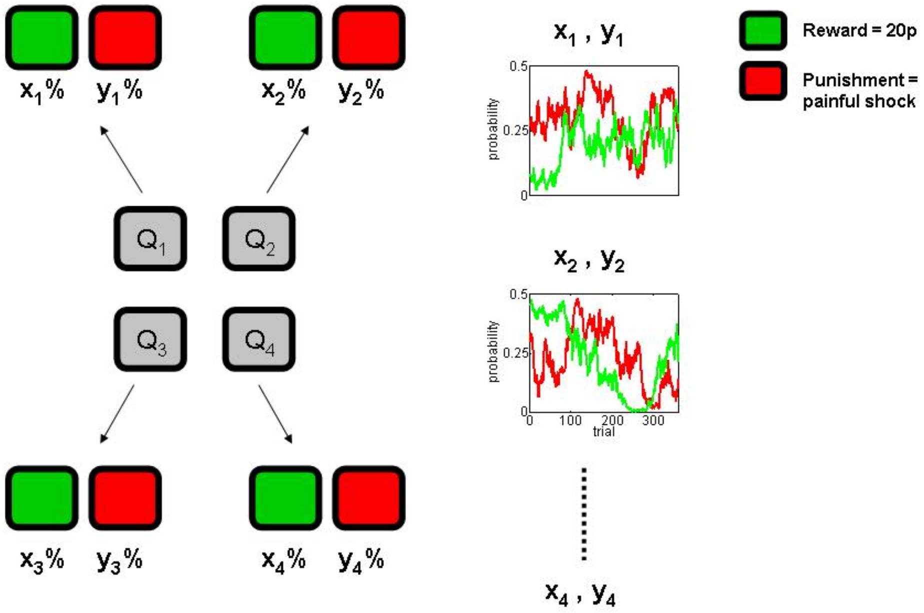


**Figure 6.1. Task design**. Subjects play a four-armed bandit task, with each bandit associated with an independent, non-stationary probability, between 0 and 0.5, of reward (20 pence) or punishment (a painful electric shock to the dorsum of the left hand). Hence subjects learn to select bandits to minimise shocks and maximise rewards, and can receive either, neither, or both on any trial.

In deciding what to choose, this task inherently requires participants to balance the values of qualitatively distinct outcomes, namely a primary aversive outcome (pain) and a secondary appetitive outcome (money). For instance, subjects could concentrate solely on winning money and ignore the pain, or concentrate on avoiding pain and ignore money, or somehow trade the two off. When different appetitive outcomes are involved, “reward prediction error” theories suggest that the neuromodulator dopamine is a candidate neural substrate for such an integrative currency(Montague et al., 2004); it is, however, less clear and indeed rather controversial whether aversive outcomes also engage dopamine or instead, another “opponent” neural system.

Importantly, this relationship between rewards and punishments relates to some of the main theories of serotoninergic function(Cools et al., 2008;Daw et al., 2002;Dayan and Huys, 2008;Doya, 2002;Robbins and Crockett, 2009). In one computationally specific version, it was proposed that serotonin serves as a motivational opponent to dopamine. However, this might have at least two effects, depending on the timescale at which serotonin opposes dopaminergic action(Daw et al., 2002). At a fast timescale, serotonin might carry an aversive prediction error, which guides aversive learning in much the same way as dopamine is thought to guide reward learning. In the context of the present task, this would predict that serotoninergic manipulation would selectively affect the impact of punishing events, by modulating how strongly or rapidly they affect behaviour, compared to rewards. Alternatively, at a slower timescale, an opponent signal might carry an average reward signal (a “comparison term” or “aspiration level”): a constant average against which individual outcomes are weighed to determine their worth. In the present task (see Methods) the effect of such a comparison would be to modulate the degree to which subjects tend to switch from the current option, or stay there, notwithstanding the outcome. Low aspiration levels lead to persistent or ‘sticky’ behaviour because, in effect, individuals are pessimistic about reward availability elsewhere in the environment. These different accounts lead to dissociable predictions of the effects of serotonin disruption in the current task.

## 6.2 Methods.

***Subjects:***

The study was approved by the Joint Ethics Committee of the Institute of Neurology and National Hospital for Neurology and Neurosurgery, and all subjects gave informed consent prior to participating. We studied 30 healthy subjects, recruited by local advertisement. We also excluded subjects according to the following criteria (numbers in brackets refer to the number of exclusions for subjects answering our initial advert).

- standard exclusion criteria for MRI scanning (2 subjects)
- any history of neurological (including any ongoing pain) or psychiatric illness (6 subjects).
- history if depression in first degree relative (6 subjects)

Female subjects were scanned mid-cycle.

***Tryptophan depletion procedure.***

We performed a randomised, placebo-controlled, double-blind, 'low-dose' tryptophan depletion procedure. This involved ingestion of a tryptophan depleted or control amino acid mix according to the concentrations below:

| | |
|---|---|
| Isoleucine | 4.2g |
| Leucine | 6.6g |
| Lysine | 4.8g |
| Methionine | 1.5g |
| Phenylalanine | 6.6g |
| Threonine | 3.0g |
| Valine | 4.8g |
| Tryptophan or placebo | 3g or 0g |

The amino acids mixture was commercially mixed and capsulated in 1g capsules (making a total of 76 capsules), and labelled according to the blinding protocol (DHP clinical). This procedure allows subjects to fully ingest all the amino acids without significant gastrointestinal side-effects, notably nausea, common with standard dose tryptophan depletion in which the mixture is prepared as a 'milk-shake'. The unblinding code was supplied in sealed envelopes opened only after the experiment had been completed.

Subjects fasted from midnight before the day of the study. On arrival on the morning of the study, blood was taken for estimation of serum amino acids. Subjects then ingested the amino acid capsules, and were allowed a small quantity of orange or apple juice (<300ml) to aid this, as well as 2-3 crackers, to ward off hypoglycaemia. Blood was taken again at 5 hrs post ingestion, just prior to the experiment. Subject timings were staggered allowing a maximum of 3 subjects to be tested per day.

To assess for side-effects as a result of the tryptophan depletion procedure, we administered a standard 10 point VAS rating scale which assesses the following criteria:

Alert / Drowsy

Calm / Excited

Strong / Feeble

Clear-Headed / Muzzy

Well-coordinated / Clumsy

Energetic / Lethargic

Contented / Discontented

Tranquil / Troubled

Quick-witted / Mentally slow

Relaxed / Tense

Attentive / Dreamy

Proficient / Incompetent

Happy / Sad

Amicable / Antagonistic

Interested / Bored

Gregarious / Withdrawn

Subjects scored higher on the aggregate side-effects at the end of the experiment (mean increase in VAS score 0.34 per item, standard error 0.21), but there was no correlation with tryp:LNAA ratio (r=-0.56, p=0.77).

We also administered the Hamilton Depression (12 question version: mood, guilt, suicide, work, retardation, agitation, anxiety (psychological and somatic), depersonalisation, paranoia, obsessiveness) before ingestion of the amino-acids, and before the task itself. This showed no evidence of pre-existing depression, and no effect on mood of the tryptophan depletion procedure.

### ***Experimental task.***

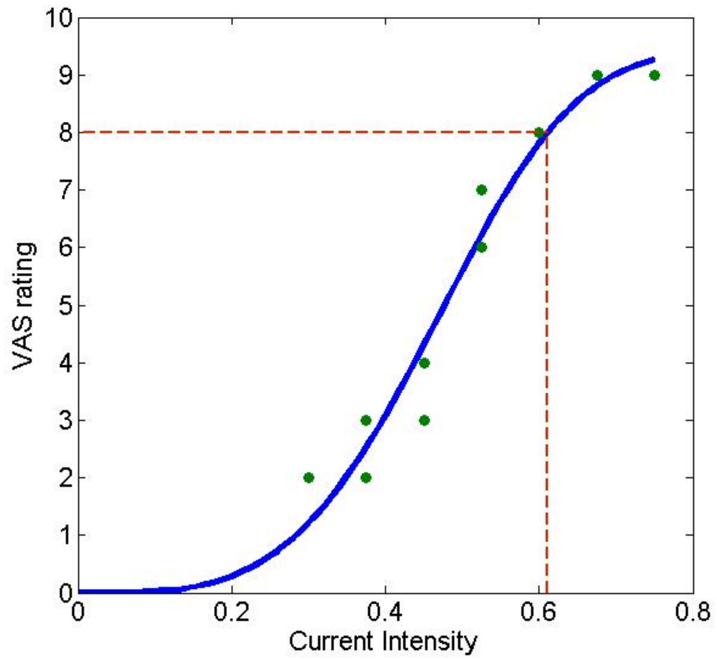
Subjects performed a probabilistic instrumental learning task involving aversive (painful electric shocks) and appetitive (financial rewards) outcomes. This equated to a 4-armed bandit decision making task, with non-stationary, independent outcomes. Each trial commenced with the presentation of the four bandits as displayed in figure 1, following which they had 3.5 seconds to make a choice. If no choice was made (which occurred either never, or very rarely across subjects), the trial would skip to the next trial automatically. After a choice was made, the chosen bandit was highlighted, and all bandits remained on the screen, and an interval of 3 seconds elapsed before presentation of the outcome. If subject won the reward, the words ‘20p’ appeared overlain on the chosen bandit. If the subject received a painful shock, the word ‘shock’ appeared overlain on the chosen bandit, and a shock was delivered to the hand (see below) simultaneously. If both a shock and reward were received, both ‘20p’ and ‘shock’ appeared overlain on the chosen bandit, one above the other, and the shock was delivered. The outcome was displayed for 1 second, after which the bandits extinguished and the screen was blank for a random interval of 1.5 to 3.5 seconds.

### ***Delivery of painful shocks***

Two silver chloride electrodes were placed on the back of the left hand, through which a brief current was delivered to cause a transitory aversive sensation, which feels increasingly painful as the current is increased. It was administered as a 1 second train of 100hz pulses of direct current, with each pulse being a 2ms square waveform, administered using a Digitimer DS3 current stimulator, which is fully certified for human and clinical use. The stimulator was housed in a aluminium shielded and fMRI compatible box within the scanner room, from which the electrode wires emerged and travelled to the subject. The equipment configuration was optimised by extensive testing to minimise RF noise artefact during stimulation.

Painful shock levels were calibrated to be appropriate for each participant. Participants received various levels of electric shocks, to determine the range of current amplitudes to use in the actual experiment. They rated each shock on a visual analogue scale of 0-10 from ‘no pain at all’ to ‘the worst possible pain’. This allows us to use subjectively comparable pain levels for each participant in the experiment.

We administered shocks, starting at extremely low intensities, and ascending in small step sizes, until they reach their maximum tolerance. No stimuli were administered above the participant’s stated tolerance level. Once maximum tolerance was reached, they received a random selection of 14 sub–tolerance shocks, which removed expectancy effects implicit in the incremental procedure. We statistically fitted a Weibull (sigmoid) function to participants’ rating for the 14 shocks and estimated the current intensities that related to a level 8/10 VAS score of pain (see below).



The participants rated another set of 14 random sub-tolerance shocks at the end of experiment, which revealed slightly lower mean ratings in the post-experimental testing, than in the pre-experimental testing procedure (mean decrement = 0.86 VAS points; standard error = 0.14). This was not correlated with tryp:LNAA ratio (r=0.07, p=0.73), showing that the tryptophan depletion procedure had no effect on pain sensation.

***Experimental procedure.***

Subjects fasted on the night before the study, and were asked to avoid high-tryptophan containing foods on the day prior to the study. They attended in the morning, consent was gained, and blood was taken for estimation of serum amino acid concentration. Subjects received a computerised tutorial explaining in detail the nature of the task, including explicit instruction on the independence of reward and punishment, the independence of each bandit from each other, and the non-stationarity of the task. Each of these points were supported by demonstrations and componential practice tasks, after which subjects moved on to perform a genuine practice task, with only the absence of shock delivery (still, however, displayed on the screen) differing from the subsequent experimental task. At this time, subjects also underwent the pain thresholding procedure. Subjects then ingested the amino-acid tablets, and relaxed in our reception area until 5hrs had elapsed, at which time blood was taken again. The subjects then entered the scanner to perform the task.

After the amino-acid ingestion, during the waiting period, subjects completed the Cloninger tridimensional personality questionnaire. Subscales for novelty-seeking (which we have previously been shown to correlate with novelty-based exploration), harm-avoidance and reward dependence did not correlate with behavioural parameters for stickiness or reward-aversion trade-off, and as such the data are not reported.

### *fMRI scanning details*

Functional brain images were acquired on a 1.5T Sonata Siemens AG (Erlangen, Germany) scanner. Subjects lay in the scanner with foam head-restraint pads to minimize any movement. Images were realigned with the first volume, normalized to a standard echo-planar imaging template, and smoothed using a 6 mm full-width at half-maximum Gaussian kernel. Realignment parameters (see analysis below) were inspected visually to identify any potential subjects with excessive head movement. This was satisfactory in all subjects, and so none were excluded.

The task was displayed on a computer monitor projected into the head coil and visible on a screen at the end of the magnet bore, visible by the subjects by way of an angle head-coil mirror. Subjects made their choices using a 4 button key-response pad held by their right side.

### *Behavioural analysis and RL model*

We used a 'direct actor' reinforcement learning model, with separate appetitive and aversive components. For instance with punishments, the learning rule is as follows:

$m^i_{reward} \rightarrow (1-\alpha_{reward})m^i_{reward} + \alpha_{reward}(o_{reward} - \rho_{reward})$ for rewarding outcomes

(1)

$m^i_{punish} \rightarrow (1-\alpha_{punish})m^i_{punish} + \alpha_{punish}(o_{punish} - \rho_{punish})$ for punishing outcomes.

(2)

And for non-chosen options:

$m^i_{reward} \rightarrow (1-\alpha_{reward})m^i_{reward}$ for rewarding outcomes

(3)

$m^i_{punish} \rightarrow (1-\alpha_{punish})m^i_{punish}$ for punishing outcomes.

(4)

Rewards and punishment action weights are integrated to provide an overall value quantity:

$$m_{\mathrm{int}} = bm_{reward} + (1-b)m_{punish}$$

(5)

Similarly, the integrated average outcome is:

$$\rho_{agg} = b\rho_{reward} + (1-b)\rho_{punish}$$

(6)

Choice is determined by the softmax choice rule:

$$p_{i=j} = \frac{\exp(\beta m_{i=j})}{\sum_i \exp(\beta m_i)}$$

(7)

For the behavioural analysis, we used a maximum likelihood method, implemented by Matlab (Mathworks inc.), to estimate the best fitting parameters of the model. Parameters were estimated (as above) on a subject by subject level, to allow us to test hypotheses relating to tryptophan status directly, and are as below:

| **Parameter** | **Depleted** | **Control** | **Contrast** |
|---|---|---|---|
| Aversive learning rate $\alpha_{reward}$ | 0.33 | 0.36 | n/s |
| Appetitive learning rate $\alpha_{punish}$ | 0.64 | 0.48 | n/s |
| Exploration coefficient $\beta$ | 16.6 | 19.9 | n/s |
| Trade-off parameter $b$ | 0.50 | 0.58 | n/s |
| Aggregate average reward $\rho_{agg}$ | -0.127, , | 0.036 | P=0.001 |

For the imaging analysis, we use the approximation $\alpha_{reward} = \alpha_{punish}$ to yield separate reward, punishment, and choice kernels, with the latter reflecting the integrated average reward term. Specifically,

$w^i_{reward} \rightarrow (1-\alpha)w^i_{reward} + \alpha o_{reward}$ for rewarding outcomes

(8)

$w^i_{punish} \rightarrow (1-\alpha)w^i_{punish} + \alpha o_{punish}$ for punishing outcomes.

(9)

For non-chosen options:

$w^i_{reward} \rightarrow (1-\alpha)w^i_{reward}$ for rewarding outcomes

(10)

$w^i_{punish} \rightarrow (1-\alpha)w^i_{punish}$ for punishing outcomes.

(11)

And a choice kernel $\chi^i$ for each option *i*

$\chi^i \rightarrow (1-\alpha)\chi^i + \chi^i$ when *i* is chosen

(12)

$\chi^i \rightarrow (1-\alpha)\chi^i$ when *i* is non-chosen

(13)

We fit a single set of parameters for all subjects, regardless of the tryptophan status, to refute the null hypothesis that there is no difference between groups.

***fMRI analysis***

Images were analyzed in an event-related manner using the general linear model, with the onsets of each outcome represented as a stick function to provide a stimulus function.

The regressors of interest were generated by convolving the stimulus function with a hemodynamic response function (HRF), and were as follows:

1. Appetitive prediction error, parametrically modeled as prediction error calculated from the reinforcement learning described in the behavioural analysis above, using the best fitting parameters at a group level (this yields more stable estimates. The prediction error was modeled at 2 time-points: the onset of the cue, and the onset of the outcome.This models rewards (money) in isolation, and ignores the aversive shocks.
2. Aversive prediction error, parametrically modeled as prediction error calculated from the reinforcement learning described above, in a similar manner to reward. This models painful shocks in isolation, and ignores the money rewards.
3. Choice kernel (stickiness value function), parametrically modeled from the reinforcement learning model above. This was modeled at the time of the cue.

Effects of no interest included:

4. Onsets of visual cues
5. Onsets of rewards
6. Onsets of the shocks
7. Realignment parameters from the image preprocessing to provide additional correction for residual subject motion.

We report activity at an uncorrected threshold in the following areas of interest, based on existing work in decision-making: ventral and dorsal striatum, medial prefrontal cortex, anterior cingulate cortex, orbitofrontal cortex, insula cortex, dorsolateral and inferior later prefrontal cortex, superior temporal sulcus, amygdala, VTA, dorsal raphe, PAG. All activities reported survive correction for multiple comparisons using 8mm sphere volumes of interest.

Note in the analysis of the choice kernel (stickiness value function), if serotonin were to negatively covary with brain activity (in the striatum) in the parametric contrast of the choice kernel, this would be consistent with its' representation of the 'missing component' of value that stems from the addition of the average reward parameter in that area. However, a positive covariation suggests that this 'missing component' must be integrated elsewhere. This is because if the propensity to stick with the current choice decreased with serotonin, then the only way that you could get a stick is to have an especially large activation in the striatum on a stick trial.

***Estimation of serum amino acid concentration***.

Immediately after venupuncture, blood was centrifuged at 3000rpm for 5mins, and serum extracted and frozen prior to analysis at -20degrees Celsius. Amino acid estimation was performed by Mike franklin, Department of Psychology, Oxford University).

***Genotyping.***

Genotyping was performed for the serotonin transporter polymorphism (SS,SL,LL alleles). The main analysis reported were tested for significant effects of allele, and allele x tryptophan status. No significant results were found, possibly because of the small number of subjects.

## 6.3 Results

Subjects performed 360 trials, concatenated over 3 sessions. We manipulated brain serotonin using a low-dose acute dietary trypotophan depletion procedure. This manipulation was between-subjects, randomised, placebo-controlled, and double-blind. Of the 30 subjects who performed the task, 15 were randomised into the control group (whose behaviour was previously illustrated) and 15 into the trypotophan depleted group. Tryptophan depletion reduces brain serotonin release, and accordingly comparison of the performance of the tryptophan depletion to control group provides insight into the function of central serotonin(Carpenter et al., 1998).

We fit subjects' choices to a reinforcement learning model (see methods). This formalises an appetitive learning process that compares phasic predictions about money to a constant average financial reward term. This is mirrored by an exactly analogous and separate aversive learning process, that learns independently about pain, and incorporates an average pain term. Choice is determined by integrating the values of each pathway, and the contribution of each is governed by an appetitive-aversive trade-off parameter.

According to the phasic opponency hypothesis, tryptophan depletion would be predicted to reduce either the punishment-reward trade-off parameter, or, alternatively, the punishment learning rate (relative to reward). This would make subjects less sensitive to punishments, or less responsive to changes in punishment contingency, respectively, and likely to accrue more pain outcomes as a result. According to a tonic opponency hypothesis, however, tryptophan depletion would be predicted to reduce the average reward signal (integrated from both appetitive and aversive learning streams), which would make subjects more persistent or 'sticky' in their behaviour.

Our data strongly support the latter hypothesis, with a significantly lower average reward term in the depleted group (-2.54 pence) compared to the control group (0.72 pence, 2-tailed t-test p= 0.001). Figure 2 shows the correlation of the average reward with pre- and post- amino-acid ingestion ratio of blood tryptophan to other large neutral amino acids, which is an accurate indicator of central serotonergic signalling. The punishment-reward trade-off parameter indicated that subjects on average judged the pain as financially equivalent to a value of 17.0 pence. However, there was no decrease in this, nor the aversive learning rate, in the depleted group. Together, these results support the suggested action of serotonin as a slow-timescale opponent to appetitive learning, carrying an aspiration level, rather than a fast-timescale opponent carrying a prediction error to drive aversive learning.

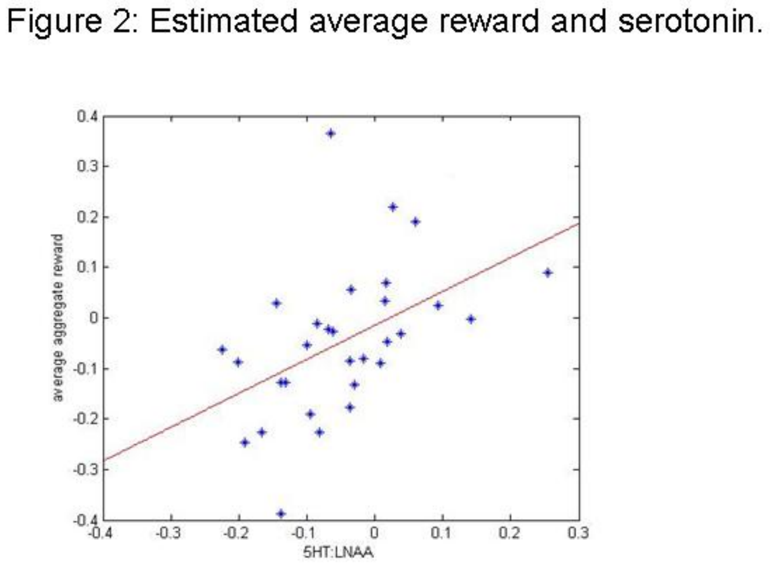


Figure 2: Estimated average reward and serotonin.

**Figure 7.2. Behavioural results: Average reward and serotonin.**

Average reward estimated from the ML fits of the behavioural data, correlated with difference between the tryptophan:LNAA ratio at the time of testing, compared to before amino acid ingestion. This measure provides an accurate index of CNS serotonin levels.

Next, we assessed brain activity correlated with the choices using a model-based fMRI approach. We used the prediction errors derived from the learning model, split into separate errors for appetitive and aversive pathways. We sought particularly to ascertain, first, whether appetitive and aversive prediction errors were integrated or separate (with the latter expected if they arise from separate, opponent systems as from dopamine and a fast, potentially serotonergic aversive opponent); whether any prediction-error-related activity was modulated by past choices as with the effect of the average reward level; and whether any of these effects were modulated by tryptophan status.

First, the appetitive prediction error was correlated with activity in striatum, as has been observed in numerous previous studies(McClure et al., 2003;O'Doherty et al., 2003) (figure 3a). Second, the aversive prediction error was *negatively* correlated with the BOLD activity in dorsal striatum (figure 3b), indeed in a region overlapping that found for positive error (figure 3c). The negative correlation with the aversive prediction error implies a positive correlation with the same signal inverted: that is, oriented like an appetitive prediction error with omitted shocks corresponding to increased BOLD activity and unexpected shocks decreased activity. In turn, this implies that the overall BOLD signal in dorsal striatum can be viewed as a single, unified prediction error in rewards minus punishments (equivalent to the difference between the appetitive and aversive prediction errors), rather than the two signals being separate. This also suggests that both effects may have a common underlying neural substrate, or, if not, that they at least converge in the dorsal striatum.

Third, we studied brain activity related to the average reward. It is not possible to directly probe a constant valued signal with fMRI, however, it is possible to look at the modulation of value induced by the average reward term on trial-by-trial choice values. With some straightforward assumptions (see methods), the average reward term is approximately manifests as a choice kernel (ie. a weight that corresponds to a tendency to repeat a choice), which adds extra value to options that have been chosen in the more recent past. For instance, a negative average reward term will produce persistence for recent choices, since such pessimism diminishes the anticipated worth of alternative options. This might be apparent as an additional value based on choice and independent of actual outcomes. This stickiness ‘value’ positively correlated with widespread activity in medial prefrontal cortex and nucleus accumbens (figure 3d).

Figure 3a: fmri results: reward prediction error

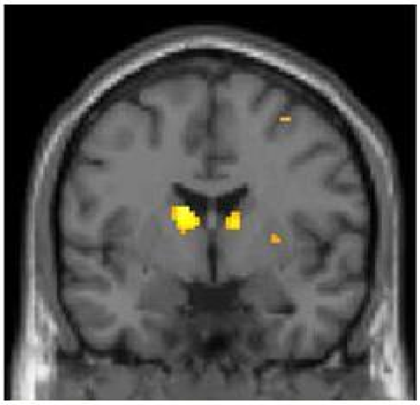


Figure 3b: fmri results: avoidance prediction error

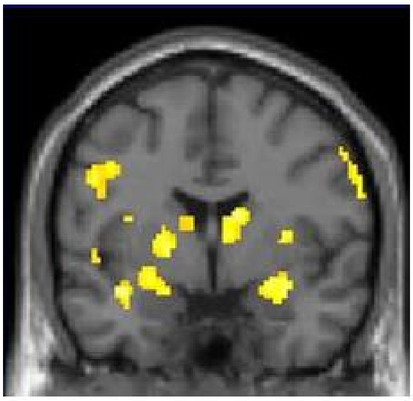


Figure 3c: fmri results: overlapping reward and avoidance prediction

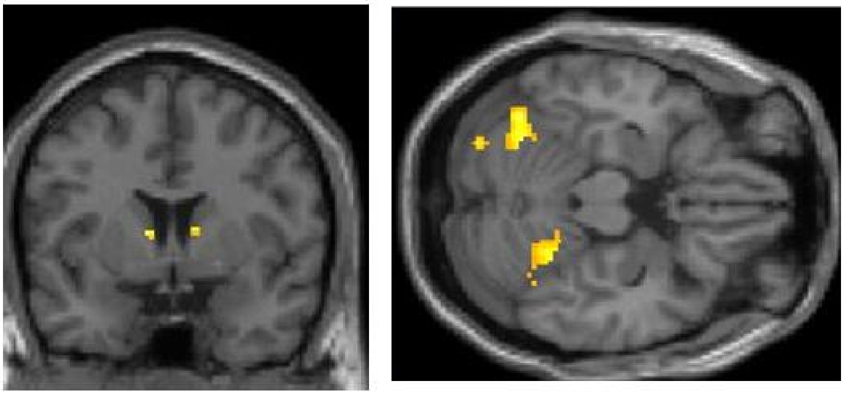


Figure 3d: fmri results: activity positively correlated to choice stickines

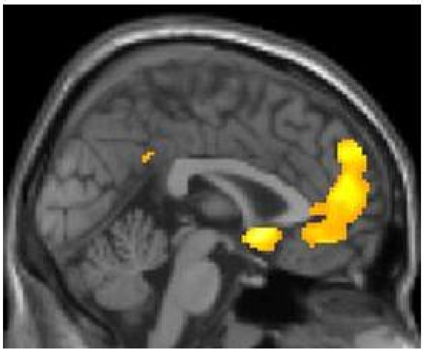


**Figure 6.3. fMRI results**

**a)** Appetitive prediction error. Bilateral head of caudate nucleus ((-12,-6,12), z=4.37; (10,2,12), z=4.29). **b)** Avoidance prediction error. Bilateral head of caudate ((14,4,18), z=4.72; (-8,2,14), z=3.35); left dorsal putamen ((-18,0,4), z=3.97); bilateral superior temporal sulcus ((-26,-2,-10), z=3.84; (34,-2,-14), z=4.34). **c)** Overlap of appetitive and aversive prediction error, showing bilateral medial head of caudate, and bilateral cerebellar cortex. **d)** Choice kernel. Activity correlating with the choice kernel: medial prefrontal cortex, nucleus accumbens.

Finally, we assessed how brain activity related to each of the above effects depended on tryptophan status. There was no significant effect on brain activity related to either the reward or punishment prediction error. This is consistent with the lack of a behavioural effect of tryptophan status on reward or punishment, and inconsistent with the hypothesis of a separate, serotonergically mediated aversive prediction error signal. There are two competing possible accounts of how the serotonin might modulate choice based on its behavioural effect on average reward. If choice value is fully constructed in the caudate, depleted subjects ought to have greater representation of a ‘stickiness’ choice kernel than control subjects. However, if the effect of serotonin is mediated outside the caudate, then depleted subjects ought to have greater representation of a ‘stickiness’ choice kernel than control subjects in this region. The data support the latter account: figure 4a shows the positive covariation between serotonin (5HT:LNAA ratio) and the stickiness choice kernel in the medial head of caudate.

**Figure 6.4. Correlation of BOLD activity and tryp:LNAA ratio**

**a)** Positive covariation between the parametric choice kernel contrast and tryp:LNAA ratio. Mdial head of caudate ((18,4,14), z=5.10; (-16,2,18), z=3.17). This shows that subjects with greater tryptophan (non-depleted) show *higher* activity in medial head of caudate associated with choice stickiness, despite their behavioural tendency to be less sticky. **b)** Correlation between tryp:LNAA and the 'stickiness' choice kernel derived from the average reward parameter in the peak voxel in right head of caudate. Note that the simple t-contrast between control minus depleted groups yields a highly similar result.

**Figure 4b: Correlation between *peak voxel* in medial head of caudate with stickiness parameter**

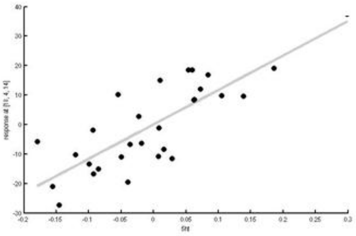
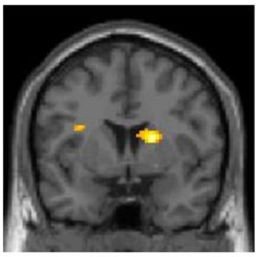


**Figure 4a: Brain activity correlated with choice stickiness**



## 6.4 Discussion

In summary, our data provide independent behavioural and neural data showing that serotonin modulates a tonic average-reward signal, that provides a comparison signal or aspiration level against which options are judged. Whereas integration of phasic opponent value prediction errors occurs in the medial head of the caudate nucleus, the data suggest that this tonic signal modulates effective value outside of the caudate.

Though behavioural "stickiness" might arise from multiple causes, one key factor, which arises in many reinforcement learning models and has previously hypothesized to be controlled by serotonin, is the effect of an average reward level (also called a comparison term or aspiration level). Such a signal provides an overall estimate of how good or bad you expect the environment to be in general, against which the individual outcomes of different choices are measured. It functions as an aspiration level in the sense that if the average reward prediction is high, then the outcomes of current options are judged marginally less attractive than if the average reward prediction is low, and as such the tendency is to switch actions and explore elsewhere, in search of higher rewards. Alternatively, if the aspiration signal is low, current options seem marginally better, so the tendency is to stick. In this way, perseveration is a direct consequence of comparing immediate vs long-term predictions. That serotonin might control long-term reward prediction has been previously predicted(Daw et al., 2002), and draws a parallel with psychological observations of serotonin's putative involvement in mood. Notably, the association of decreased serotonin signalling with depression offers at the very least a phenomenological link to the notion of reduced aspirations about future reward.

There are other factors that may also contribute to choice stickiness, though these have not previously been linked to serotonergic function. For instance, it could result from a simple form of stimulus-response (SR) learning, in which previously taken choices are 'stamped-in'(Mackintosh, 1983). Or it can be viewed as a process to encourage oversampling of information. This latter process may be advantageous in widely variable environments during which reinforcement learning has a tendency to be over-sensitive to the immediate past, which can lead to risk-aversion(Denrell and March, 2001).

Although the lack of an observed effect of tryptophan depletion on aversive learning or aversive-appetitive opponency does not exclude such a role for serotonin, the comparison with the magnitude of the choice effect is particularly striking. A caveat to the presumptive refutation of these theories is the persistent uncertainty about exactly what aspect of serotonin signalling is disrupted by tryptophan depletion. This uncertainty extends to broad anatomical differences (subcortical versus cortical), timescale differences (phasic versus tonic) and synaptic dynamics (direct signalling versus autoregulation). This suggests the need to explore different methodologies of serotonin function in future studies, such as neuronal recordings, cellular imaging, and psychopharmacological studies.

The data also refine the role of the striatum in motivation. Previous Pavlovian punishment studies (in which punishments are delivered regardless of any action) have shown an aversive prediction error signal, oriented positively (opposite that seen in the present study) in the ventral and dorsal striatum(Jensen et al., 2006;Seymour et al., 2004), suggesting a site of convergence with the (putatively dopaminergic) reward prediction error. However, in the present study, the sign of the aversive signal changes to a reward-signed signal. The key difference between the studies may be the availability of choices in the present design. If so, this would be consistent with "two-factor' theories of instrumental avoidance, in which avoidance is mediated by the "reward" of attaining a "safety state" that indicates successful avoidance (Dinsmoor, 2001;Mowrer, 1947). The comparison of these studies suggests that in passive studies on aversion, punishments may be framed as punishments, but when control becomes possible, punishments may be framed as missed appetitive opportunities of avoidance. In fact, this is consistent with previous demonstrations of reference sensitivity of striatal activity, in which the contextual valence is apparently set by predictive cues (Seymour et al., 2005). Critically, by forcing independent representation of reward and avoidance, our data show that avoidance prediction, carried as an opponent reward-predictive signal, co-activates the same region of striatum (medial head of caudate) as that involved in signalling the prediction of standard rewards. This demonstrates that the head of caudate is an integration site for these distinct motivational pathways. Whereas this appetitive-aversive integration (algorithmically, the addition of appropriately scaled excitatory and inhibitory values (Dickinson and Dearing MF, 1979;Mackintosh, 1983)) is commonplace in everyday decision tasks, this is possibly the first direct experimental demonstration of its neuroanatomical basis.

However, the data also yield a surprising result with respect to the role of the striatum in choice. In particular, the data argue against the values expressed (by way of error terms) in the striatum as being the sole determinant of choice, given the anti-correlation of caudate activity with perseverative behaviour. Rather, it suggests that striatal value processing must be integrated with an average reward signal elsewhere. As mentioned above, locating the anatomical substrate for such a tonic signal may be difficult with fMRI, because the temporal frequency of noise in fMRI acquisition may be similar to that of a slowly varying average-reward signal. However other methodologies, such as lesion data, may be more informative. Of particular  note,  selective prefrontal (and not striatal) serotonin lesion studies in the marmoset monkey lead to inflexible, perseverative choice, suggesting that this may be the mediate of an average-reward serotonergic effect on choice(Clarke et al., 2004;Clarke et al., 2007).

Lastly, the task provides a novel way to determine the aversiveness of incommensurable quantities such as pain. Judgements of pain have typically relied on explicit ratings (in humans), or innate responses (in animals). The limitations of these methods are well established, in particular for human rating scales which are at the mercy of a range of subjective influences(Fields HL and Price DD, 2005). The task permits assessment of judgements of aversiveness without explicit ratings, based instead on (economic) choice.

# Chapter 7. Discussion: contributions.

In this chapter, we highlight some of the specific novel contributions the research has made to the field. The subsequent chapter (‘The architecture of aversive motivation’) assimilates our findings with others in the field to put forward an integrated theoretical structure of aversive motivation, including appetitive-aversive integration. Lastly, the final thesis chapter (‘Implications for related disciplines’) provides a supplementary discussion of the implications of our and other findings have towards related disciplines, in particular behavioural economics and social neuroscience.

## 7.1 Methodological contributions:

### *7.1.1 Developing computational models of human pain and aversive behaviour*

Psychological and neurophysiological approaches to human pain have generally been phenomenological, orientated around explicit (reportable) and implicit responses that can be measured and categorized. Hence, inferences about the underlying physiology have been reverse, driven by the structure of the observable responses. Accordingly, the dominant theory of the sub-structure of central nervous system pain systems is a tri-fold dissociation of ‘sensory-discriminative’, ‘cognitive-evaluative’, and ‘affective-motivational’, based in no small way on introspective evaluation of what pain ‘feels like’. Here, however, we have taken the opposite approach, concentrating on what function pain evolved to perform, and proposed a *generative* model which is tested by its ability to reproduce behaviour and predict neural responses. The thesis represents the first formalised attempt to approach pain in this way.

Our goal here has been to study the motivational function of pain – how the brain learns to predict pain, and how these predictions shape behaviour. Although motivation and decision making is well studied in human pain science, there have been remarkably few attempts to incorporate the theory and findings of animal learning literature. Rather, the existing standard draws on studies of traditional human psychology.  Our thesis is primarily based on the framework which animal learning theory provides, and illustrates the value of this approach by its success in permitting simple, testable models of behaviour.

More generally, the thesis represents an engineering approach to pain and pain motivation, by proposing formal mathematical models of behaviour that yield quantitative predictions. The animal learning theoretic framework is formalised within computational models, which are proposed *de novo* here. The reinforcement learning framework espoused in the initial theoretical work has strong parallels with that independently developed in studies of reward learning (Schultz, Dayan, Montague 1999).

### *7.1.2 Applying model based fMRI to ask novel questions about brain mechanisms.*

Computational models of pain prediction can yield quantitative models of behavioural responses in experimental tasks. In instrumental learning, this is evident by the choices that the test subject makes, which is typically easily defined (for instance, in a forced binary option paradigm). In Pavlovian learning, the response is often less easy to discriminate (for instance, pupil dilatation or skin conductance). However, even if it is, it may not be sufficient to use such responses to test competing hypotheses about the computational structure of processes that yield these responses. In this way, neurophysiological data can provide adjuvant evidence to support the validity of different models, which may be critical if competing models provide different predictions regarding the mechanistic processes involved in generating the ultimate behavioural output. In the case of prediction learning, one such process is the generation of prediction errors.

Our approach yields two ways in which neurophysiological data can be informative: firstly, it can provide evidence to support a computational model by showing that the subcomponents are represented (anywhere) in the brain. Secondly, it can identify where in the brain a component process is represented, and so allow such findings to be incorporated with the body of neuroscience data that underlies the general understanding of the role of different brain areas in pain and motivational learning.

The methodology that we use to achieve this is model-based fMRI, and uses linear regression of parameters derived from a (hypothetical) computational model computed on a trial-by-trial basis. This approach was developed in our lab initially to study Pavlovian reward learning (O'Doherty et al, 2003), and used to show that existing theory and data from primate reward learning experiments could also predict neural responses in human reward learning. The data in this thesis represent the first application of this approach to test a fundamentally new computational theory (ie. one that was not developed elsewhere in other experimental domains).

## 7.2 Computational and psychological contributions:

### *7.2.1 The validity of TD models for Pavlovian aversive learning*

Although the thesis set's out to formalise the motivational basis of pain, the theory generalises, and is generalised, to any aversive outcome (such as financial loss). The novel theoretical framework proposed is based on a view of aversive outcomes as quantities to be minimised, in the context of an agent that can learn about its (uncertain) environment through interacting with it. The core idea in the thesis formalises this in terms of the Bellman equation, and proposes Reinforcement Learning algorithms as plausible ways in which the brain can solve the problem (ie. of predicting and minimising pain).

As mentioned above, Pavlovian responses themselves (pupillary diameter, reaction times, skin conductance) are of insufficient fidelity to track the subtle acquisition of learning on a trial-by-trial basis, and this reinforces the attempt to use fMRI to provide further evidence. Ultimately, the data that support of the temporal difference model of Palvovian learning are striking: the correlation of BOLD responses with the temporal difference prediction error is remarkably robust. This finding has been replicated in both our subsequent studies, and by other authors, such that the reinforcement learning (TD) model of aversive motivation has become a widely accepted theory within the field.

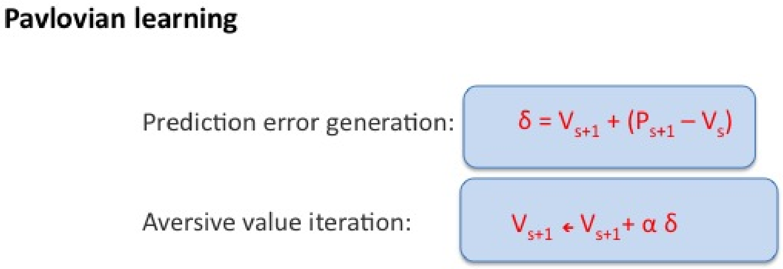

Pavlovian learning
Prediction error generation:
$\delta = V_{s+1} + (P_{s+1} - V_s)$
Aversive value iteration:
$V_{s+1} \leftarrow V_{s+1} + \alpha \delta$

Figure 7.1. Pavlovian learning: $V_s$ and $V_{s+1}$ represent the aversive value at successive states. Punishment P is delivered on state transition, and δ represents the prediction error, which is used to update the state value, to an extent dependent on the learning rate α.

### 7.2.2 *Extension of TD models for avoidance learning.*

As we discuss in more detail in the next chapter, instrumental learning in the face of aversive outcomes (escape and avoidance) is more complex than Pavlovian learning, since ultimately the emitted behaviour involves the coordination of both Pavlovian and instrumental processes. However, it is still possible to treat the instrumental component as a single process, in a Thorndikian manner. Here, we formalise this in much the same way as in the Pavlovian case, in terms of error-based value learning rules.

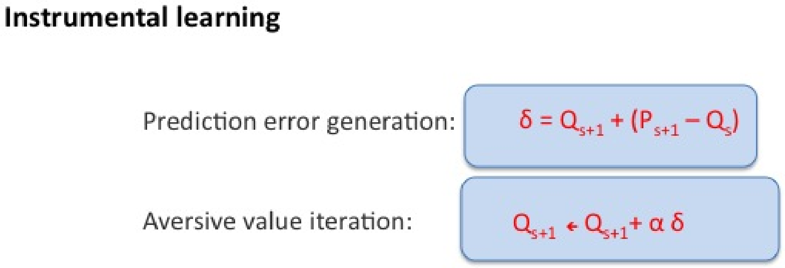

Instrumental learning
Prediction error generation:
$\delta = Q_{s+1} + (P_{s+1} - Q_s)$
Aversive value iteration:
$Q_{s+1} \leftarrow Q_{s+1} + \alpha \delta$

Figure 7.2. Instrumental learning: $Q_s$ and $Q_{s+1}$ represent the aversive action value at successive states. Punishment P is delivered after taking an action, and δ represents the prediction error, which is used to update the action value, to an extent dependent on the learning rate α.

Within reinforcement learning, there is in fact a number of different ways of implementing trial-and-error action learning. One can either learn the true expected values associated with different actions, and then choose amongst the available actions based on these values (an ‘indirect actor’). Alternatively, one can iteratively learn action ‘weights’ directly, without going via the calculation of expected values (a ‘direct actor’ method) (Dayan and Abbott, 2001). However, both share the same general reliance on an action-based prediction error term to guide either value or policy iteration, respectively. Our data finds robust evidence for this error term, and provides compelling evidence for the validity of Reinforcement Learning models of instrumental avoidance learning. This is illustrated by the models ability to predict subjects’ actual choices on a trial-by-trial basis.

### *7.2.3 The existence of opponent motivational systems*

One of the awkward facts about neural information coding is that neurons cannot fire both positively and negatively to encode a full scale of positive and negative quantities. Indeed, the only way that neurons can achieve this is to have a tonic baseline firing rate, and to encode negative quantities by pauses or reduction in that baseline. But clearly this seems an inadequate way to deal with aversive values, especially given the potentially important (eg. life threatening) nature of the outcomes they convey.

The notion of distinct appetitive and aversive motivational systems has existed for some time in experimental psychology, and indeed the notion of single opponent, mutually inhibitory systems is supported by a number of ingenious experiments in animals (Dickenson and Dearing, 1979). Here, we formalise this in terms of opponent temporal difference processes, and show that this implemented in a mirror opponent manner (as opposed, for example, to a rectified opponent) in the brain. This represents the first directly evidence-based computational account of dual appetitive and aversive motivation in humans.

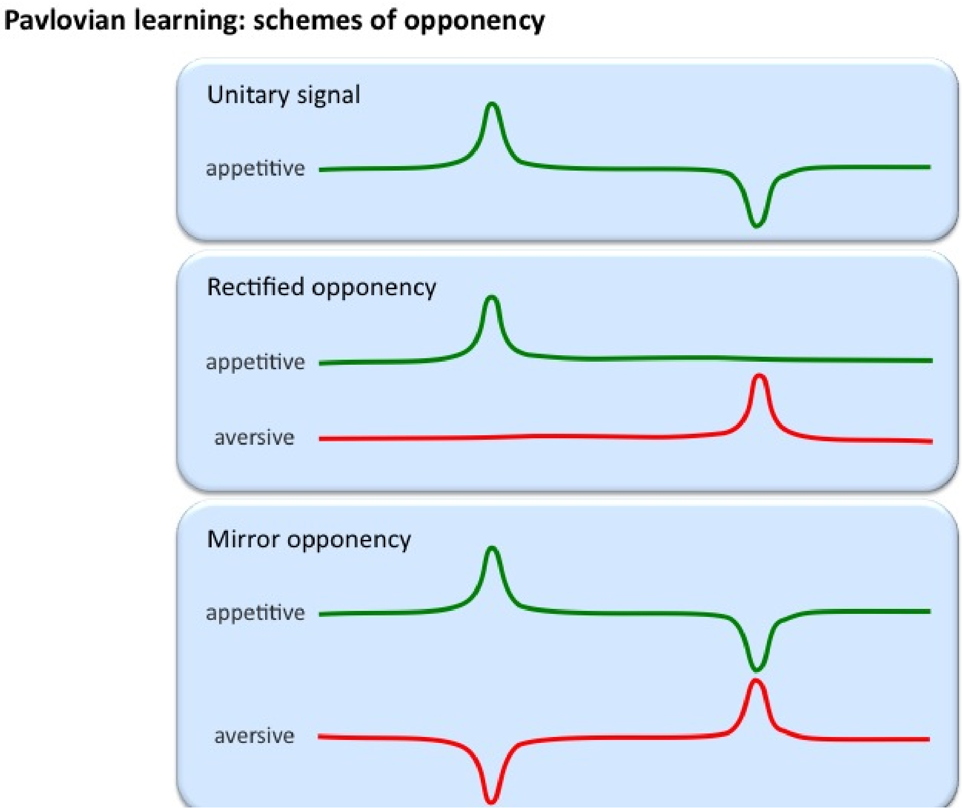


Figure 7.3 Different possible schemes of opponency. The neurobiological data support mirror opponency, implement in amygdala and putamen (appetitive system) and lateral orbitofrontal cortex and putamen (aversive system).

### *7.2.4 Integrated choice model.*

Especially important is an understanding of how these opponent systems are integrated to provide a unified metric to guide choice. Our thesis describes three important processes within this, given representations derived from separate independent opponent streams. First, the brain must generate an opponent appetitive representation of punishment that acts as a ‘safety signal’ to guide successful avoidance. Second, appetitive and avoidance representations must be appropriately scaled in magnitude. Third, the brain must summate the values of each. We show that this is implemented in the brain, and identify a unified action value error term. Accordingly, our thesis provides a basic account of integrated appetitive and aversive motivation across both Palvovian and instrumental learning.

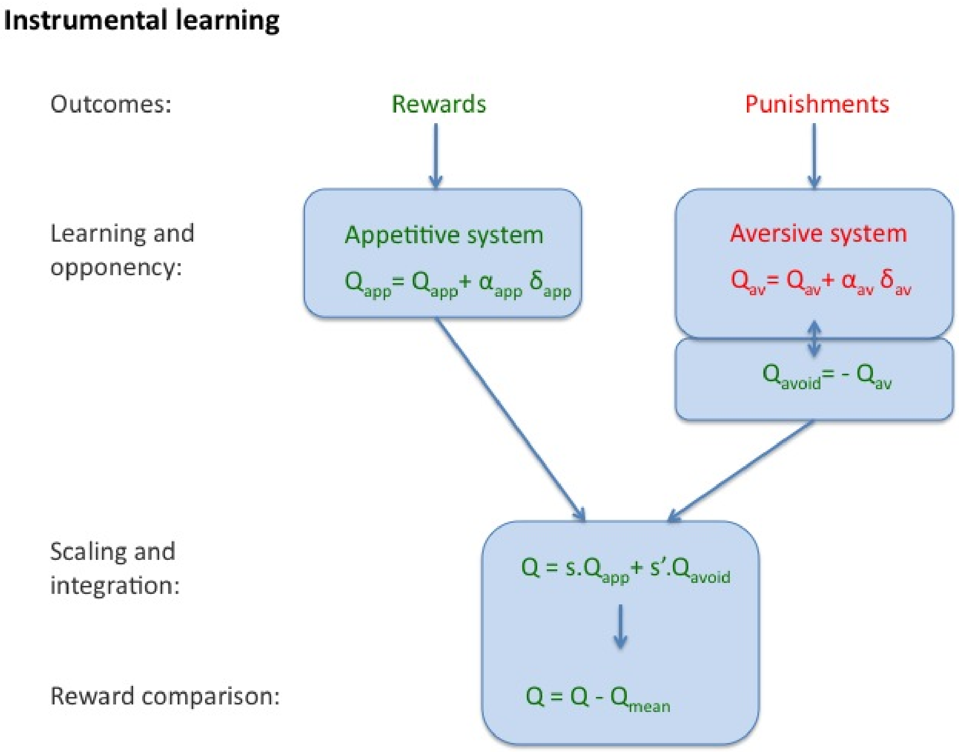


Figure 7.4 Different stages in integrated instrumental choice. Separate appetitive and aversive systems compute action values for rewards and punishments respectively. The outputs of each are scaled and summated, and related to a general comparison term (average reward signal).

### *7.2.5 Average reward models.*

We add one further complexity to our Reinforcement Learning account of motivation, namely the representation of average-reward (or punishment). This emerges in the Pavlovian case, from administration of tonic punishment, in which the amount of tonic pain acts as a reference point from which perturbations are subsequently judged. In the instrumental case, average reward models provide a putative account of the perseveration of choice, independent of actual outcomes, that is typically witnessed in tasks in both humans and primates. Its implementation sees values and actions compared to an average expected quantity that acts as a sort of ‘aspiration level’, rather than in absolute terms.

In both Pavlovian and instrumental cases, the evidence of the representation of average reward is indirect since it is not possible to directly observe (ie. image) the tonic outcome signal. However, it does provide the simplest and most parsimonious account of the data.

## 7.3 Neurobiological contributions:

### *7.3.1 The role of the basal ganglia*

One of the key findings of the work has been uncovering the role of basal ganglia structures in aversive motivational learning. This has had particular impact since the human neuroimaging field at the time largely viewed structures such as the striatum and substantia nigra as reward specific. Indeed, observations of activity in these regions in complex tasks typically led to the reverse inference that the task recruited appetitive motivational pathways. In contrast, the amygdala was subject to a similarly widespread (mis)conception as a structure involved almost exclusively in aversive motivation. Despite the fact that both these accounts were clearly questionable after even briefest review of the animal learning literature, they were undoubtedly widely held. As such, our findings have played an important role in changing our understanding of the role of the human basal ganglia in motivation.

At the heart of this has been the data that has shown that bilateral ventral putamen encodes an aversive temporal difference error. This is manifest in our studies of electrical pain, thermal pain, and financial loss, suggesting that it represents a common aversive motivational process. Furthermore, we have shown that this is implemented as a fully signed error signal – in that it codes positive and negative values with increased and decreased BOLD activity respectively. The activity co-localises with activity seen in comparable studies of reward learning, which suggests the anatomical integration of motivational learning systems within the ventral putamen. We also show evidence of an anatomical dissociation along an anterior-posterior within the putamen, with more aversive specific activity localising to posterior puamen, and appetitive specific activity localising to more anterior-ventral putamen, towards the nucleus accumbens. However a large region of mid-ventral putamen appears to be sensitive to both aversive and appetitive motivational prediction errors.
Our first imaging study (chapter 3) also identified aversive prediction error activity in caudate and substantia nigra, indicating that this pathway is expressed

more widely than just the ventral putamen. The activity in caudate is notable, since anterior / head of caudate prediction error activity has also been observed in instrumental conditioning tasks, including ours in the final experimental chapter. Previous studies of both Pavlovian and instrumental conditioning have suggested dissociation between ventral putamen and caudate activity in Pavlovian and instrumental tasks respectively (O'Doherty et al, 2004). The paradigms adopted in our research have been exclusively either Pavlovian or instrumental, but it is clear that instrumental effects may exist in Pavlovian designs, and Pavlovian effects may exist in instrumental designs. Unless one uses both instrumental and yoked Pavlovian designs and compare the two, which we have not done here, it is difficult to make strong inferences about anatomical specificity Pavlovian or instrumental systems within our data.

That being said, our final instrumental study identifies solely the head of caudate in the dual representation of simple appetitive and avoidance errors. On the basis of previous studies, therefore, it seems highly likely that this represents an instrumental prediction error signal.  What is most interesting is the anatomical superposition of the two (simple appetitive and avoidance) error signals, which arise from independent outcome statistics. Although this does not necessarily imply functional integration, since it could still be feasible for the systems to be distinct at a neural level, it does seem likely that this activity may play a role in motivational learning that integrates reward and punishment. As such, both ventral putamen and head of caudate emerge from our data as probable key brain regions in the integration of appetitive and aversive motivational learning.
The data do not exclude the possibility of appetitive-aversive integration elsewhere. In particular, our paradigms are designed to optimally identify prediction error related activity, and not the representation of the aversive and appetitive values themselves. It is likely that other brain regions may do this, in particular the orbitofrontal cortex (for Pavlovian values) and ventromedial prefrontal cortex (for instrumental values).

### *7.3.2 The anatomy of opponent systems*

Our data also illustrate the anatomy of opponent motivational systems outside of the basal ganglia. In particular, we find evidence for aversive prediction error representation in the lateral orbitofrontal cortex, and appetitive prediction error representation in the amygdala.

The nature by which each is part of a functionally connected motivational system, for example between lateral OFC and basal ganglia, remains unclear. Correlated activated between distant neural regions could be driven by a single neuromodulator, or could be driven by functional cortical-basal ganglia-cortical loops. A further difficulty is in knowing whether the activity represented represents synaptic activity (ie afferent input), or neuronal activity, or both. Thus BOLD correlates might represent serial connectivity in a functional pathway. For instance, prediction errors might be expressed in basal ganglia (for instance, via a neuromodulator), which mediates the storage of aversive *values* via cortical-basal ganglia-cortical loops in lateral OFC.

Lastly, we note that evidence of the role of the amygdala in relief provides an especially limpid demonstration of this region’s role in appetitive motivation. This is especially striking given the nature of this representation in the absence of any primary rewards. That is, the representation is purely inhibitory, reflected either termination of tonic pain, or omission of expected phasic pain. This illustrates the spectrum of opponency within a nucleus which has been at the heart of studies in emotion and motivation (which we have discussed elsewhere).

### *7.3.3 The role of serotonin.*

Finally, our data offers a new perspective on the function of serotonin. Undoubtedly, the diversity of projections and receptor subtypes has complicated the search for general theories of serotonin function, but it remains likely that within this complexity may be a computationally specific representation that is of value to a diverse range of neural functions. Using an appropriately sophisticated decision task, our data illustrate a remarkably selective effect of tryptophan depletion on choice (in both behavioural and fMRI data), in which it controls a component of choice flexibility independent of immediate outcomes. This suggests that it might be a slow timescale average reward signal, acting as a comparison term or aspiration level in decision-making. In this way, greater release of serotonin signals greater ‘hope’ about available rewards in the environment, against which immediate outcomes are judged. This in turn leads to greater flexibility and exploration (and less ‘pessimistic perseveration’), which provides an intriguing phenomenological parallel with conventional accounts of the role serotonin in mood.

In summary, the thesis provides the first computational account of aversive motivational learning in humans, with the basal ganglia at the heart of its neurobiological implementation. In the next chapter, we assimilate these findings with existing data in the field to propose a basic general account of aversive motivational systems.

# Chapter 8. Discussion: the architecture of aversive motivation

Below, we integrate the results presented in the previous chapters with other data to provide a general overview of the neurobiology of aversive learning. This centres on the mechanistic structure of aversive learning in humans. We discuss the accumulated behavioural and neurobiological evidence for multiple value systems, and show how they are exploited by distinct action systems to allow a range of aversive behaviours to emerge. Given that aversive learning has evolved from one traditionally considered to have the amygdala at its heart, we pay special attention to this region's emerging more general role in affective decision making, and we highlight its role in Pavlovian-instrumental interactions.

## 8.1 Value systems.

Aversive control requires some method of valuing both actual and predicted losses. Understanding the different mechanisms by which this is achievable draws on the computational problem of how this value is acquired in an uncertain environment. That the world consists of naturally beneficial and threatening outcomes has inspired theoretical models, most notably reinforcement learning, that learn how to evaluate and act in the world based on experience, and learn online using trial and error. Insights from reinforcement learning have been remarkably successful in illuminating the neural mechanisms of motivation and decision making, not least since some of the algorithmic solutions of the general reinforcement learning problem seem to have direct neural implementations. Below we describe the different aversive value systems and their neural bases.

### 8.1.1 Innate values

Certain stimuli are endowed with an inherent aversiveness. Pain, for instance, is subserved by a sophisticated system of specialised nociceptive pathways signalling of actual or imminent tissue damage to many areas of the spinal cord and brain (Hunt and Mantyh, 2001). This results not just in a set of characteristic, often involuntary, defensive responses, but also a perceptual representation of negative hedonic quality. This illustrates the innate affective impact that reflects the evolutionary acquisition of value, guided by generations of reproductive success.

In humans, innate aversion is often accompanied by conscious experience. Indeed, the feeling associated with loss dictates the way these systems are often described in traditional psychological accounts. This can be approached more formally by considering ‘feeling’ as a process of hedonic inference. As with many less motivationally loaded sensory systems, afferent information is rarely perfect, and a statistically informed approach is to integrate afferent input with either concomitant information from other modalities (multi-sensory integration), or prior knowledge of events (expectation).

In the brain, the basic representation of innate value implicates brainstem and midbrain structures, including the amygdala, periaqueductal grey, parabrachial nucleus, and thalamus. Cortical structures such as insula are associated with aversive representations across modalities, including conscious negative hedonic experience (Craig, 2002).

### 8.1.2 Forward-model values

The immediacy of innate values renders them poor at guiding more planned decisions, and undoubtedly the explicit anticipation of losses has an important role in shaping decisions. Naturally, control systems should optimally exploit value systems that involve prediction of an aversive event before it occurs, since it allows possible escape or avoidance of it. One way of doing this is to generate a hypothetical (‘imagined’) representation of an anticipated loss, incorporating some sort of model of the state changes that might take you there. This sort of forward-modelled value system is a key part of what might traditionally be regarded as a cognitive value system, in that, in humans at least, they seem to draw on an explicit representation of a future event.

The offline evaluation of aversive value, in which sequences of future events can be ‘run-through’ in abstract representation, and values corresponding to intermediate events calculated, bears similarity to dynamic programming methods in reinforcement learning. Such iterative valuation schemes consider putative distal goals and punishments and try to inferentially work out the value of more proximal states. Forward modelled aversive events are perhaps the least well understood of all value systems, by virtue of their necessary complexity.

One of the remarkable, if slightly informal observations from Pavlovian conditioning experiments (including those in chapters 3-5) is that the majority of subjects are not aware not only of the true contingencies, but of the very existence of contingencies at all. This suggests that such forward-modelled values, insofar as they might be expected to be available to awareness, may in some circumstances be inferior to other (cached, see below) value systems in picking up statistically viable aversive contingencies in the environment.

Some additional insight into the dissociation, both behavioural and neurobiological, between forward-modelled and cached values comes from a recent experiment in which we explicitly sought conscious, contingency awareness during Pavlovian conditioning, drawing on the observation that contingency awareness interacts with conditioning differently across different acquisition schemes (Clark and Squire, 1998;Han et al., 2003;Knuttinen et al., 2001;Ohman and Soares, 1998). Specifically, successful trace conditioning (in which there is a temporal gap between the offset of the CS and onset of US) is thought to be more dependent on explicit awareness, suggesting that perhaps these values are more related to some form of goal representation. In the study (Carter et al., 2006), we simultaneously conditioned human subjects to predict an aversive electrical stimulus (US) from arbitrary visual cues (CS) with concurrent delay and trace protocols: to assess contingency awareness, subjects reported their shock expectancy on each trial, and we also recorded skin conductance as a putatively more implicit measure of conditioning, to identify conditioning that *wasn't* under conscious awareness. Our data indicated a clear role for the middle frontal gyrus in contingency awareness during conditioning, correlated specifically with the acquisition of awareness on a trial-by-trial basis. This was contrasted with amygdala activity, which reflected acquisition of implicit knowledge, as indexed by autonomic activity.

### 8.1.3 Cached values.

In many real-life decision problems, anticipating precisely when and where aversive outcomes may occur becomes difficult. Three things contribute to this: i) sequentiality, in which outcomes depend on long trains of actions or state changes, ii) stochasticity, whereby outcomes are uncertain, either with known (risk) or unknown (ambiguous) probabilities, and iii) non-stationarity, in which probabilities drift over-time, either slowly or abruptly.

One way round at least some of this complexity is to collapse the total anticipated value of future state transitions or actions on those that are immediate. This can be termed caching, in honour of its relation to a similar process in computer science. In effect, a cached value provides a single metric as to the overall utility of a particular state, or taking a certain action. It integrates over the uncertainties of the various outcomes, and the times when they might be expected, to report how bad (or good) it is.

Reducing much of the complexity of the future onto a single value is clearly attractive, not least since it considerably simplifies action control, as we discuss later. What had been less obvious, at least initially, is how an individual has access to such a value. Our evidence indicates that the brain follows the simple algorithmic scheme described by temporal difference learning (chapter 3). This method prescribes an experience-based way of continually refining cached value estimates, using discrepancies (prediction errors) between adjacent estimates. Using sequential estimates to transfer value between adjacent states, as opposed to waiting for outcomes themselves (as in Monte Carlo methods, for instance (Sutton and Barto, 1998)), provides an effective way of propagating value to states more distant from an outcome. However, the computational simplicity comes at the cost, in comparison to forward modelled values, of efficiency, since updating is tied to experience.

From an implementational perspective, the aversive temporal difference error is expressed clearly in the striatum, across different modalities of aversion, and in part-overlapping / part-distinct (more posterior) regions of striatum (chapter 5).

The cached values themselves (which are equivalent to ‘expected values’ in economics) are represented in anterior insula.

Further support for this latter finding comes from ERP data, based on the design presented of the pilot study described in Chapter 1. Source localisation of high density EEG recordings of anticipatory activity in pain prediction shows temporally precise predictive value representations localising to anterior insula (Brown et al., 2008b;Brown et al., 2008a).

### 8.1.4 Long-run average values

One of the deficiencies of phasic cached values is that it tells you little about the distribution of punishments (or rewards) over time. Furthermore, many aspects of behaviour benefit from a temporally more broad perspective than that tied tightly to individual actions and states. Although behavioural experimentalists often require strictly cue-evoked responses, the natural environment is rarely so precise. Consequently, estimating a diffuse, temporally integrated average value may be a valuable quantity.

Theoretically, average values might be used in several respects. First, since there are almost always costs tied to actions and responses, they can be used to determine the overall rate of responding that optimising returns. Second, they can be used to make broad judgements to guide exploratory behaviour, that is, drive exploration when short term cached values are lower than long-run average values. Third, in hierarchically structured environments or decision processes, they can be useful in valuing higher level states and actions.

Long-run average aversive values are best tied to cues that share their temporal outlook, and hence are naturally aligned to contextual information. More phenomenologically, their representation may have a bearing on mood states (including physiological stress and depression), and have a natural correspondence with tonic primary aversive stimuli such as chronic pain. In the brain, the methodological difficulties in tracking slowly changing neurophysiological responses over extended times (in contrast to phasic, cue-evoked responses) make the current knowledge about representation more uncertain. However, their existence can be inferred indirectly from experiments that look at phasic perturbations of tonic stimuli. Accordingly, we have shown that relief of tonic pain, as an aversive inhibitor, elicits a positively signed appetitive prediction error in striatum, in contrast to a negatively signed aversive prediction error (chapter 4). Furthermore, we see striking predictive activity in lateral orbitofrontal cortex, which hints (in the absence of any formal demonstration) that this region may be specifically involved in computing phasic value in the context of tonic value. This latter suggestion would certainly be in keeping with other data on orbitofrontal cortex.

Secondly, we have also shown in chapter 6, behavioural and neurally, albeit indirectly, that average outcome models provide the best fit for data when it comes to aspiration and exploration. Critically, modulation of this level, that is the putative interaction between tonic and phasic predictions) implicated the medial head of caudate, overlapping with the representation of phasic action value prediction errors.

Third, this tonic outcome representation appears to be modulated by serotonin, in a manner consistent with previous suggestions that serotonin mediates a tonic aversive outcome signal. This is notable given the long-standing association of serotonin with depression, for which low reward aspiration might be a plausible underlying computational component.

## 8.2 Control systems

The different value systems outlined above play distinct roles in guiding actions in the face of aversive events. As we discuss below, there is good evidence that control is governed by several distinct action systems that relate closely to the different value systems. Ultimately, aversive events need to be escaped from, reduced or avoided if possible, and each of these behaviours draws on different controllers in specific, and occasionally complex, ways.

### 8.2.1 Goal directed control

Goal-directed actions are characterised by the existence of some sort of representation of the outcome of an action, relating closely to the representation of forward-modelled values. In animals, this is well illustrated in aversive devaluation experiments. In this, an animal is first trained to perform an action for a food reward. Next, the food is separately paired with experimentally induced nausea and vomiting. When subsequently tested on the original action, animals often perform it much less often than the appropriate controls, suggesting that they have constructed some form of internal representation that the action leads to the ill-effects.

In humans, goal-directed action is closely affiliated to ‘cognitive’ control, in which individuals explicitly consider the outcome of actions, and of subsequent actions, and use some form of tree-search to inform current actions. The brain might support different ways of doing this, for instances using propositional, linguistic structures, or spatially based structures. It has affinity with the classical notion of outcome-expectancy expounded by Tolman (Tolman, 1932), and with more recent fields such as dynamic programming in engineering.

Although substantial regions of prefrontal cortex may be involved in goal-directed control, the ventromedial prefrontal cortex appears to be fairly central to goal representation (Ariely and Norton, 2007b). Rat lesion experiments have indicated that this region exploits connections with dorsomedial striatum. In aversive goal-orientated control, the prefrontal cortex is likely to involved as well, although this has yet to be clearly shown.

### 8.2.2 Habitual control

Habits relate strongly to cached values, learned through trial-and-error. They lack any representation of the outcome or subsequent available actions that result from taking action. Instead, they represent only the utility of the action itself.

Habits rest critically on the state (discriminative stimulus) to inform whether and which habits are available, and thus appear to be stimulus driven.

If the environment is relatively stable (stationary), then habit-learning provides a near optimal strategy for selecting actions. However, caching, by its very nature, can take a long time to learn, especially in complex environments. Furthermore, any type of rapid change in the environment cannot flexibly be accommodated. Thus in these situations, and in the context of limited experience, goal-directed control may be superior.

As mentioned above, dopamine projections, particularly from substantia nigra to the dorsolateral striatum, are crucially involved in learning appetitive habits. Dopamine neurons are thought to modulate plasticity in cortico-thalamic loops which ultimately store habits. We have shown previously the role of striatum in both simple instrumental appetitive and avoidance action (note, we did not explicitly differentiate goal-directed and habit systems), and that a region of the striatum appears to treat aversive inhibition (avoidance) indistinguishably from appetitive excitation (reinforcement) (Pessiglione et al., 2006).

In chapter 6, we showed that integrated decision making involves separable but convergent learning systems. Because the study forced independence of rewards from punishments, the representation of avoidance errors was necessarily distinct from simple reward reinforcement errors. The requirement to make one decision at one time forced subjects to integrate these values, trading off the independent valence magnitudes of each. We showed that a simple Direct Actor reinforcement learning well describes both subjects’ behaviour, and their neurophysiological (BOLD) responses. We found that the dorsal striatum (medial head of caudate) is critically implemented in this, suggesting it is a critical site for appetitive-aversive integration in action control.

### 8.2.3 Pavlovian control.

Innate values typically have a set of characteristic responses associated with them. Often these are primitive, such as increased heart rate and sweating during acute pain, or fighting in the midst of a contest. Such responses are evolutionarily appropriate actions, and appear to be hard-wired into the brain. Pavlovian learning provides a natural extension of this, by eliciting responses to stimuli that reliably predict innately salient events. Thus, not only does contingency between neutral stimuli and intrinsic rewards and punishers (unconditioned stimuli) engender the acquisition of a cached Pavlovian value, it also elicits a response appropriate to it. However, such responses are not simply duplicates of those produced by the conditioned responses themselves (stimulus substitution), but typically carefully anticipate the event they predict.

Pavlovian responses fall into two behaviourally and neurobiologically distinct types. ‘Preparatory’ responses reflect the general valence of the predicted outcome, and elicit non-specific responses such as approach or withdrawal. ‘Consummatory’ responses reflect the specific attributes of the outcome, such as salivating and licking for foods, and leg flexion for foot-shock. Indeed, in the aversive domain, it appears that the repertoire of consummatory responses is both complex and sophisticated, arguably much more so than in the appetitive domain. Classically, defensive responses have been divided into fight, flight or freeze, although the precise nature of the response is both varied, and depends rather precisely on the nature of the outcome and the context in which it is predicted. For instance in rats, anticipation of a shock causes freezing if the cue is generalised, attempted escape if it is localised, fighting in the presence of another male, and copulation in the presence of a female. Clearly, the specificity of these responses has been carefully moulded by evolution, and indeed the exact nature of the responses is often highly species specific. But most notably, they interact with other control systems in important ways.

In the brain, Pavlovian responses have been well studied. The acquisition of aversive Pavlovian values depends most critically on the amygdala. The central nucleus is predominantly involved in directing non-specific preparatory responses, including arousal and autonomic responses and approach/withdrawal, which is achieved by projections to various brainstem nuclei, including reticular formation and autonomic nuclei. The basolateral complex is predominantly involved in much more specific, consummatory response, mediated downstream through connections to regions such as the hypothalamus and periaqueductal grey. The latter has a sophisticated, topographically organised architecture mediating the range of defensive to aggressive behaviours.

We explored the aversive role of the PAG in a simple, ecologically inspired maze-task, in which subjects were chased around a maze by a computerised predator, analogous to the 'ghosts' in the classic 1980's arcade game 'Pac-man' (Mobbs et al., 2007;Mobbs et al., 2009). These predators, however, administered either an electric shock if they caught the subjects before the end of the (variable duration) trial. We found that the PAG encoded the interaction between predatory imminence and predator magnitude, and furthermore, this predicted subject-specific scores on threat-susceptibility behaviour on a psychological questionnaire. In a follow up study, we sought more direct evidence of Pavlovian actions (Mobbs et al., 2009). Using variable intensity punishment (painful electric shocks) to signify capture, we looked at occasional panic-like responses that occur when capture is imminent. These were correlated with PAG activity, and suggest the intrusion of impulsive escape responses over skilled avoidance. Future work is planned to explore the modulatory role of serotonin in this paradigm, to test the hypothesis that 5HT inhibits panic, but increases anxiety (Graeff, 2004)

## 8.3 Constructing aversive behaviour.

The evidence of multiple control systems raises the question of whether they act independently (competitively) or together (cooperatively) in guiding aversive behaviour. As we show below, most behaviours involve cooperative integration of the different systems, such that the very existence of co-acting systems is often superficially obscure. It takes instances of more direct competition, often involving indictment of the Pavlovian system, to betray the different strategies of the underlying control systems.

### **8.3.1 Pavlovian-instrumental interactions**.

The simplest illustrations of punishment can be observed by attaching aversive contingencies to actions pre-trained with rewards. For instance, if a rat has learned to press a lever to receive a food pellet, then replacing the food pellet with an electric shock causes the animal to press the lever less often, and indeed stop pressing it all together. A number of early experiments established that this effect was sensitive to basic statistical and economic manipulations. First, aversive outcomes of higher intensity have a greater inhibitory effect on actions. Second, aversive outcomes that follow actions with greater certainty are more effective in suppressing action. And third, aversive outcomes that occur more imminently are more potent. This latter effect illustrates the basic phenomenon of temporal discounting, which as for rewards, declines the magnitude of events as they become less imminent.

Importantly, these basic suppressive effects reflect more than one process. At first glance, they would appear to reflect basic habit-based or goal-orientated action reduction. However, a number of early experiments had difficulty in showing any instrumental component at all, with a wealth of data implicating Pavlovian mechanisms. One of the reasons for this is the nature of aversive Pavlovian responses causes them to be appropriate in very many situations, which is a testament to their sophistication. However, appropriately controlled experiments (employing for instance, yoked Pavlovian-instrumental designs) illustrated that instrumental contingencies clearly enhance the suppressive effect of aversive outcomes. Furthermore, the Pavlovian component operates in two ways. First is the direct contribution of the Pavlovian action: for example withdrawal starts to become incompatible with pressing a fixed lever. Second, the Pavlovian value itself suppresses the action by a phenomenon called conditioned suppression. The latter process is illustrated by the fact that merely presenting a Pavlovian cue during instrumental responding for a reward, suppresses responding.

Conditioned suppression is the mirror image of the appetitive phenomenon of Pavlovian-instrumental transfer (PIT). One of the key features of appetitive PIT is that it is composed of at least two dissociable components. The first is a non-specific process by which appetitive Pavlovian conditioned stimuli excites appetitive actions non-selectively. This depends on the integrity of the central amygdala and nucleus accumbens shell (Cardinal et al, 2002). The second component is a specific process by which conditioned stimuli selectively augment actions towards outcomes with which they are associated. This depends on the integrity of the basolateral amygdala and nucleus accumbens shell. Conditioned suppression, as the aversive equivalent of PIT, is necessarily a non-specific appetitive-aversive interaction, and has been shown to depend on the central amygdala.

Rationalising conditioned suppression in a theoretical framework can draw on two aspects. First, the non-specific nature of the behavioural suppression naturally absorbs any uncertainty as to whether there is indeed a specific contingency between an action and an aversive outcome. In any decision theoretic framework, this reflects a ‘safety-first’ approach that makes economic sense. Second, particularly in the context of long-run average values, cues can be thought of as influencing (reducing) an overall assessment of average expected return. In the face of opportunity costs, this relative value ought to reduce the rate of responding.

Another illustration of Pavlovian instrumental cooperation occurs in conditioned punishment, which differs from conditioned suppression in that the separate Pavlovian and instrumental values are integrated more in series, than in parallel. In conditioned punishment, an individual will learn to perform an action less often if it results in presentation of an aversive Pavlovian cue. This mirrors conditioned reinforcement in the appetitive case, and provides an important illustration of how Pavlovian values can be used as surrogate goals to suppress or reinforce instrumental actions. The result reflects the integration of the cached values of each.

### 8.3.2 Avoidance.

At the heart of aversive control is avoidance. Clearly, the goal of behaviour is to learn to avoid aversive events wherever possible. However, consideration of the problems that must be solved in avoidance hint, quite correctly, that such behaviour may not be straightforward. For instance, how are successful avoidance actions reinforced, if by definition they lead to no outcome? How does an individual ever realise that the threat is gone, if never sampled? Understanding how the brain solves these problems is crucial, but requires a fairly close look at the experiments that have engaged animal learning theorists for many decades.

In a typical avoidance paradigm, an experimental animal receives a warning cue (such as tone or light), that precedes delivery of an aversive stimulus (signalled avoidance), such as prolonged electrification of the floor of the compartment. At first, the animal responds only during the aversive stimulus, and successfully escapes if it jumps into a neighbouring compartment. After several presentations, the escape response is executed more quickly, and eventually, the animal learns to jump when observing the warning cue, thus completely avoiding the shock.

Mowrer was the first to formally assert that learning to avoid involved two processes (Mowrer, 1947): first was to predict the threat, and second to learn to escape from the predictor. These processes, proposed respectively to be under Pavlovian and instrumental control, comprise two-factor theory, which in one form or another has survived well over the past decades. Although there are many unanswered questions about precisely how the different action systems are orchestrated in different avoidance situations, some key facts are well grounded.

Notably, Pavlovian mechanisms play a critical (and multifarious) role in avoidance, and indeed Pavlovian responses to the warning cue (the discriminative stimulus) alone are often capable of executing successful avoidance (Dayan and Seymour, 2008). For example, jumping out of an electrified chamber, blinking in anticipation of an eye-puff, leg flexion to an electric foot-plate can all completely remove an aversive stimulus, without any need for an instrumental component. That they do pays tribute to their evolutionary provenance, and led some to question the involvement of instrumental responses at all.

Several experiments demonstrate the role of the Pavlovian cue. For example, presenting a separately trained aversive cue during avoidance increases avoidance responding (a form of Pavlovian-instrumental transfer). Furthermore, animals will learn to avoid a cue that has been independently pre-trained with an aversive stimulus.

The importance of the instrumental contingency is demonstrated by the fact that some avoidance responses such as lever-pressing, key-pecking (for pigeons) are difficult to reconcile as aversive Pavlovian responses. That they are much harder to train than some other responses suggests that avoidance responses may be executed over a basis set of Pavlovian actions. Furthermore adding instrumental contingencies to yoked Pavlovian avoidance designs improves avoidance.

However, whereas this delineates a role for instrumental escape, it fails to yield any role for the avoided state, which is typically signalled if only by the termination of the warning cue. Indeed, avoidance is impaired if termination is delayed, and improved by presentation of additional cues that signal successful avoidance. Indeed, such cues have been shown to reinforce separate avoidance responses.

These results are consistent with the notion that the value of a safety state following successful avoidance reflects a Pavlovian aversive inhibitor. Importantly, such values share a common representation with appetitive excitatory values, demonstrated by their ability to block them (transreinforcer blocking). That this state plays an important role in control is suggested by the fact that avoidance responses continue long after the Pavlovian aversive responses to the discriminative stimulus have extinguished (as they will of course do if avoidance is successful). Thus it may be more than circumstantial that in purely Pavlovian designs, conditioned inhibitory values are somewhat resistant to extinction.

This places the role of the Pavlovian value attached to the discriminatory stimulus in the spotlight (Bersh and Lambert, 1975;Biederman, 1968;De Villiers, 1974;Kamin et al., 1963;Mineka and Gino, 1980;Overmier et al., 1971a;Overmier et al., 1971b;Starr and Mineka, 1977), since on the one hand it ought to act so as to suppress instrumental actions that lead to the aversive outcome, and on the other hand it ought to encourage instrumental actions that lead to the appetitive safety state. But there is more to avoidance than just the classical contingency: animals can be trained to perform one response in the presence of one discriminative stimulus and a different response to avoid the same shock in the presence of a different stimulus. Avoidance warning stimuli can suppress appetitive instrumental behaviour, in a similar fashion to conditioned suppression by an aversive CS, but this effect is diminished with prolonged expression of the avoidance response. This effect, as Starr and Mineka showed in a classic experiment (Starr and Mineka, 1977), is over and above the effect of classical extinction due to the repeated success of avoidance. What seems clear therefore is that what is required to establish a successful avoidance response in not necessarily the same as what is required to maintain it

The dissociation of components in avoidance is supported by neural data. Selective lesions of central or basolateral amygdala impair conditioned suppression, and conditioned punishment selectively (Parkinson et al, 2000). Neuroleptics interfere with learning avoidance responses, but not acquisition of instrumental escape responses (Cook and Catania, 1964). In human studies, in support of the role of appetitive pathways, dorsal striatum and ventromedial prefrontal cortex display reward-signed activities during avoidance. Furthermore, they do so in a manner predicted by reinforcement learning models (chapter 6). However, what is currently lacking is selective lesions that dissociate goal-directed and habit-based components of the avoidance action. The existence of a goal-directed component is illustrated by sensitivity to outcome in experiments that manipulate body temperature in the context of avoidance actions that lead hot or cold outcomes, which are differentially appetitive or aversive according to body temperature (Henderson and Graham, 1979). Beyond that, however, it has not been very thoroughly studied. Furthermore, few animal studies have explored how avoidance values are integrated or dissociable from appetitive reinforcement values.

## 8.4 The role of the amygdala in motivation and learning

In human imaging neuroscience, the prevalent view has been that the amygdala is the predominant seat of aversive learning. Indeed fMRI studies have suggested both that aversive Pavlovian values are acquired, and prediction errors expressed, in amygdala in a dynamic fashion consistent with prediction error based models (Glascher and Buchel, 2005a;Yacubian et al., 2006). Temporal prediction errors, which encode discrepancies between both predictors and outcomes (embodied in reinforcement learning models such as temporal difference learning), have been observed in ventral striatum (Jensen et al., 2006;Seymour et al., 2004), but not amygdala. This raises the question as the precise role of the amygdala in aversive (and appetitive) motivation.

In monkeys, lateral habenula neurons provide an aversive signal that inhibits dopaminergic neurons during negative reward prediction errors (Matsumoto and Hikosaka, 2007). In the amygdala, single neuron recording studies have identified neurons that encode the Pavlovian value of rewards, punishments, as well as neurons that encode salient, valence independent predictions (Paton et al., 2006). A recent study of probabilistic appetitive and aversive conditioning has shown that separate neuronal populations encode valence specific, probabilistic value-related signals (ie. modulated by outcome uncertainty). Furthermore, some neurons showed evidence pointing towards a mirrored opponent pattern of activity, in which they coded both reward and omitted punishment, and vice versa. This suggests that amygdala neuron learning might be driven by a temporal prediction error signal (no cells intrinsically displayed a full prediction error pattern themselves) arising from elsewhere.

How these values are acquired is not yet clear. In theoretical models of Pavlovian learning, learning is often thought to be guided by a prediction error, which updates values based on the discrepancy between predicted and actual outcomes. For appetitive values, this is thought to be guided by dopaminergic projections from the ventral tegmental area in the midbrain, particularly to ventral striatum (Nakahara et al., 2004;Satoh et al., 2003;Schultz et al., 1997). However, whether dopamine directly ‘teaches’ neurons in the amygdala, or alternatively some other mechanism such as transfer of values via connections from the ventral striatum, is not clear. In the aversive case, a comparable neuromodulator to dopamine has yet to be discovered, although as we discussed in the preceding chapter, 5HT has been suggested (Daw et al., 2002).

The functional impact of negative prediction errors in the aversive domain has theoretical importance, since omission of aversive stimuli guides extinction learning. Aversive extinction is appetitive in valence, just as omission of appetitive stimuli is aversive (and can block primary aversive stimuli (Dickinson and Dearing MF, 1979)). This (aversive extinction) is known to be mediated by active learning that involves inputs from medial PFC (Maren and Quirk, 2004;Milad and Quirk, 2002). Critically, extinction memories are easily 'forgotten' or disrupted by procedures such as reinstatement, and are sensitive to reconsolidation (Duvarci et al., 2006). This aversively biased asymmetry endows amygdala based Pavlovian values with the same sort of 'safety-first' encoding that reflects the affective hard-wiring of unconditioned stimuli. Thus it is possible that the temporal difference based mechanisms of Pavlovian value learning in striatum reflect a more flexible and distinct alternative system to that implemented in amygdala, even though both use prediction errors.

So what is the broader role of the amygdala in learning and motivation? A number of studies illustrate the distinct roles of CE and BLA in mediating Pavlovian-instrumental interactions. For instance, Killcross and colleagues took rats with either CE or BLA lesions, first trained them in a Pavlovian conditioning procedure, and subsequently tested them in an instrumental procedure in which actions led to presentation of the CS (Killcross et al., 1997). CE lesioned animals displayed a deficit in the non-specific suppression of instrumental responding (conditioned suppression) produced by the CS, whereas BLA lesioned animals exhibited a deficit in biasing instrumental choices away from an action that produced the CS (conditioned punishment). In another key experiment, Corbit and Balleine, using a selective satiation procedure for instrumental actions that lead to different rewards, demonstrated that CE lesions (previously implicated in PIT (Hall et al., 2001;Holland and Gallagher, 2003)) selectively impaired general forms of PIT, but that specific forms were selectively impaired with BLA lesions (Corbit and Balleine, 2005).

The dissociable roles of the CE and BLA have been shown in many other Pavlovian and Pavlovian-instrumental tasks. In addition to general PIT and conditioned suppression, the BLA appears to be critical for contextual conditioning (Selden et al., 1991), conditioned approach (Hitchcott and Phillips, 1998) and conditioned orienting (Holland et al., 2002a) . Furthermore, beyond mediating specific PIT and conditioned punishment (as part of avoidance), the BLA is critical for reinforcer revaluation (Balleine et al., 2003;Hatfield et al., 1996;Malkova et al., 1997), conditioned reinforcement (Cador et al., 1989;Hitchcott and Phillips, 1998) and second-order conditioning depend on BLA (Burns et al., 1993;Hatfield et al., 1996).

These results suggest that the BLA encodes specific value-related outcome information, such as that modulated by satiety. Some of anatomical connections that subserve this are suggested by a series of elegant experiments on conditioned potentiation of feeding. In this paradigm, Pavlovian cues paired with food when individuals were hungry are able to motivate sated animals to eat beyond satiety. Rats with lesions of the BLA, but not CE, do not show the characteristic potentiation of feeding normally seen when the Pavlovian cues are presented (Holland et al., 2001;Holland et al., 2002b). This effect depends on connectivity with hypothalamus and OMPFC (Petrovich et al., 2002;Petrovich et al., 2005), but not striatum or lateral OFC (McDannald et al., 2005). Indeed, a wealth of other experiments have confirmed the importance of amygdala-OFC connections in mediating the impact of outcome-specific value representations on choice (Baxter et al., 2000;Baxter and Browning, 2007;Ostlund and Balleine, 2007;Paton et al., 2006;Saddoris et al., 2005;Schoenbaum et al., 2003;Stalnaker et al., 2007)

Amygdala connectivity with nucleus accumbens mediates a number of Pavlovian influences on action. Firstly, autoshaping (and also higher order conditioned approach), which reflects Pavlovian actions, depends on the integrity of BLA, accumbens, and connections between them (B.Setlow et al., 2000;Parkinson et al., 2000;Parkinson et al., 2002). This may be an important mediator of the Pavlovian impulsivity seen in paradigms such as negative automaintenance (Dayan et al., 2006;Williams and Williams, 1969). Second, lesions of the core and shell of the accumbens disrupt specific and general forms of PIT, respectively (Corbit et al., 2001).

Amygdala connectivity with prefrontal cortex may mediate more outcome specific influences on action. More specifically, the medial prefrontal cortex (prelimbic cortex in rats) is critical for learning action-outcome contingencies (Balleine and Dickinson, 1998;Bechara et al., 2000;Hampton et al., 2006;Kim et al., 2006). Disrupting connections between BLA and mPFC impairs avoidance choice in conditioned punishment (Coutureau et al., 2000).

The role of amygdala in humans has been highlighted in the context of patients with amygdala damage (Bechara et al., 1999), who like patients with ventromedial prefrontal damage, are impaired decision making tasks involving risk and uncertainty. However, the pattern of impairments differs in that amygdala patients have clear deficits in Pavlovian processes. Hampton and colleagues recently showed that patients with amygdala lesions showed (using fMRI) impaired outcome representations of instrumental choices in ventromedial prefrontal cortex (Hampton et al., 2007).

Indeed, many of the animal results have strong parallels with human experiments (Delgado et al., 2006;Phelps and LeDoux, 2005). Amygdala and OFC are both implicated in specific representations of outcome value in a similar manner to animals (Gottfried et al., 2003). The role of this circuit in controlling decisions may underlie many aspects of human behavioural economics. For example, amygdala and OFC are involved in using previous experiences of regret to bias future decisions (regret avoidance)(Coricelli et al., 2005). Amygdala activity also reflects the interaction of emotionally framed information with risk-based option choices, for instance motivating risk aversion in positive contexts (De Martino et al., 2006). Furthermore, the relative aversion of humans to ambiguity, as compared to risk, is linked to activity in the amygdala activity (Hsu et al., 2005).

These latter studies points to the possible importance of the amygdala in risk and uncertainty, which has interesting, though speculative, links with experiments in rats (Gallagher and Holland, 1994). Notably, lesions of the CE appear to impair the increase in learning due to increases in CS-US uncertainty (Holland and Gallagher, 1993). Associability is theoretically aligned to ambiguity by the fact that both drive learning, in contrast to risk. The control of learning by the former heavily implicates the neuromodulators acetylcholine and norepinephrine, midbrain sources of which (nucleus basalis and locus coeruleus, respectively) both receive substantial input from the CE.

To summarise, distinct regions of the amygdala appear to play a critical role in modulating decision making. Thus the CE may play a critical role in optimising metalearning, both through outcome non-specific modulation of approach and rate of responding possibly via dopaminergic modulation of ventral striatum, and rate of learning through acetylcholinergic modulation of more diffuse cortical areas. In contrast, the BLA may have a more specific role in optimising choice, utilizing refined outcome specific knowledge gained from connections with hypothalamus and OFC, and via projections to goal-specific areas, in particular the ventromedial PFC (infralimbic cortex in rats).

To conclude, we advance the viewpoint that the amygdala is not just involved in Pavlovian conditioning with the goal of executing simple conditioned responses, but is especially concerned with integrating Pavlovian values with habit based and goal orientated systems, across both aversive and appetitive motivation, mediated principally via connections with striatum, and ventral and orbitomedial prefrontal cortex respectively.

# Chapter 9 Discussion: consequences for behavioural economics.

## 9.1. Historical and methodological issues.

Traditionally, emotion has been embedded within a two-system model of human decision-making, a conceptual framework still dominant in psychology and behavioural economics. In its simplest form, it reduces to a deliberative, cognitive system viewed as a ‘cold’, rational and far-sighted, operating alongside an affective, emotional system which is ‘hot’, irrational and short-sighted (Camerer et al., 2005;Kahneman and Frederick, ;Sloman, 1996) . Although this structure provides a very effective descriptive tool across a diversity of situations, the extent to which it can encompass emergent empirical neurobiological findings in decision making is increasingly doubtful. Indeed, we have described above the extensive evidence from both animals and humans that illustrates the probable operation of multiple decision-systems. Furthermore, it may be that the processes that mediate emotional influences on decisions (which are likely to frequently be Pavlovian) are often rational, and it is just they are often only *apparent* in instances when they are not.

We first mention a couple of methodological points about the relationship between economic and psychological paradigms. In behavioural economics, decisions are often probed in relation to options with stated parameters, that is, the magnitudes, risks and uncertainties of various options are given directly. These are likely to exert their effects mostly through model-based predictions (and goal-directed control). By contrast, in experimental psychology, the parameters of options are typically learned through trial and error. Thus, representations of value and risk are experience-based rather than propositional, and can have an impact through model-free as well as model-based control. Of course, experience-based representations are imperative in animal experiments, and have also been highly successful in deconstructing the components of aversive (and appetitive) behaviour. However, any complete account of aversive behaviour needs to integrate both, since humans are presented with both types of situation: one off decisions such as those regarding pensions and life insurance; and repeated decisions, such as those regarding what painkiller to take or which foods to buy.

A further difference in methodologies relates to type of aversive events used. Whereas economists usually use monetary loss, neuroscientists have often used more diverse, primary stimuli such as pain: for instance in the form of an electric shock to hand or paw. The advantage of this is it is an immediately and relatively instantaneously consumed commodity. Furthermore, it is both potent and ecologically valid, in the sense that it is the sort of stimulus with which aversive systems evolved to deal.

An important distinction, across the different classes of value discussed above, is that between excitatory and inhibitory values. Inhibitory values arise from the opponent relationship between aversive and appetitive events, and the nature of the relationship between the two is well studied in animals. Inhibitory aversive values arise when either appetitive events are omitted, or when tonically received appetitive stimuli cease. Importantly, there is a natural consistency between appetitive inhibitors and aversive excitators. For example, in terms of value representations, omission of food is intrinsically similar to painful shocks (demonstrable in psychological paradigms such as summation and blocking). Likewise, there is a natural opposition between aversive excitators and aversive inhibitors (again, demonstrable in retardation and counter-conditioning paradigms).

The naturally opposite relationship between appetitive and aversive values is also evident when one considers their physiological function. For example, hunger and thirst (beyond the typical physiological range) can be aversive, and signal lack of a reward (indeed, excessive food may even be aversive). Likewise, financial loss is identical to a lack of financial reward. This raises an issue, with important theoretical consequences, that echoes through both psychological and economic accounts of loss behaviour, namely on the distinction between homeostasis and heterostasis. Homeostasis is a feature of physiological systems, in which motivation is directed to maintain, or restore physiological equilibrium.

Heterostasis reflect motives that are monotonically increasing. The vast majority of motives are homeostatic, and this important consequences for the predicted shape of utility functions. In simple terms, aversive events move away from homeostatic equilibrium, and rewards move towards it. However, whether this usefully explains all motives is doubtful. Rewards such as sex may be a special case of heterostatis because of the genetic and evolutionary consequences, and non-perishable commodities such as money (and storable food in some species), may buy long-term homeostatic stability, and thus be effectively heterostatic.

## 9.2 Pavlovian influences on economic choice.

Insight into the importance of Pavlovian mechanisms can be gained for considering the type of information, both specific and general, that Pavlovian values carry. In the general sense, Pavlovian states represent an estimate of the expected value of being in a particular state, and thus cues may provide an indication of the average amount of reinforcement available at a given time. This turns out to be a potentially very useful signal. First, it provides a standard against which individual actions should be judged: for example, receiving £5 is positive in a neutral context, but negative in the context of Pavlovian cues that inform that the average outcome is £10. Not only does this change the relative utilities of available options for individuals with non-linear utility functions across positive and negative outcomes, but relative judgments may influence exploration and apparent risk attitudes if the value of outcomes has to be learned (Denrell, 2007;March, 1996;Niv et al., 2002). This is because if the value of an action is uncertain, then the relative value of an outcome determines the frequency with which it is sampled: an option judged aversive will be tried less often than one judged positive. Thus, Pavlovian values can modify the asymmetrical sampling biases between positive and negative, or high versus low variance outcomes.

Second, in addition to judgements of *relative* utility, Pavlovian values can also usefully inform how much effort an individual should invest in a set of actions.

This notion embodies the concepts of excitement and motivational vigour, and can be rationalised in any system in which there is an inherent cost to performing an action (Niv et al., 2007). If the average return is judged high by a Pavlovian system, then it makes sense to invest more effort in instrumental actions, as seen in general Pavlovian-instrumental transfer. In this way, emotional values mediated by the Pavlovian system are integrated, synergistically, with other action systems in a way that exploits the distinct information embedded therein.

Third, and more specifically, Pavlovian values can selectively alter the value of different options presented simultaneously. Pavlovian cue value reflects a state-based homeostatic quantity which reflects physiological need: for example the utility of food declines as one becomes sated, or the utility of shelter is reduced on a fine, warm day. This information can be used to judge the specific utilities in situations in which many courses of action exist, as is demonstrated by sensory-specific satiety. Indeed, one of the paradigms (devaluation) that has been particularly instructive in dissociating different action systems draws on the fact that habit-based learning systems are unable to access specific value related information without experiencing outcomes and relearning actions (Balleine, 1992).

### 9.2.1 Impulsivity.

Impulsivity covers a broad range of phenomena. Classically, it features engagement in actions whose immediate benefits are less than those of longer term pay-offs that would accrue if the subjects could be patient (Cardinal et al., 2004). That is, subjects exhibit temporal short-sightedness. Impulsivity is best described in the appetitive domain, but similar notions may apply in aversive domains too. In the appetitive case, we have argued that the effect of a Pavlovian approach response associated with a proximally available beneficial outcome can be to boost early, and thus impulsive, responding at the expense of what would be favoured by goal-directed or habitual instrumental systems (Dayan et al., 2006). Treating this form of impulsivity in Pavlovian terms amounts to a subtly different explanation of the behaviour from accounts appealing to (or data fitting with) hyperbolic discounting  or indeed ideas about differences between (model-based) rational and (model-free or perhaps neuromodulator-based) emotional cognition, which conventionally ignore the normative intent of model-free control. In the aversive domain, impulsivity may be manifest as intrusions of innate aggression in the face of loss. Later, we suggest that one route to altruistic punishment is via Pavlovian aggression.

### 9.2.2. Framing effects.

Framing effects are a rather well-studied peculiarity of human choice in which the decision between options is influenced by subtle features of the way in which those options are presented. Typically, the language used to describe an option is manipulated in a valance related manner, whilst the expected value remains unchanged. This biases choices in a reliable manner, and violates a central tenet of rational decision-making, namely logical consistency across decisions, regardless of the manner in which available choices are presented. This assumption, known as 'extensionality' (*1*) or 'invariance' (*2*), is a fundamental axiom of Game Theory (*3*). However, the proposition that human decisions are "description-invariant" is challenged by a wealth of empirical data (*4,5*). Kahneman and Tversky originally described this deviation from rational decision-making, which they termed the "framing effect", as a key aspect of Prospect Theory (*6*, *7*).

In a well-known example of this, experienced physicians were asked to recommend optimal management (surgery or radiotherapy) for a hypothetical cancer patient. Remarkably, they advised radically different treatments depending on whether the treatment information had been presented in terms of either mortality or survival rates (*5*). Another well known example is the disease dilemma, in which subjects are asked to choose between two options relating to the management plan of an epidemic, one of which contains risk, and the other not (Tversky and Kahneman, 1981b). The risky option is fixed, such as 'Option A has 2/3 chance of curing all 600 affected people', but the non-risky option is presented in either a positive or negative frame, as either 'With Option B, 200 people will be *saved* ' or 'With Option B, 400 people will *die*'. Subjects are more likely to choose the risky option when the sure option is presented in aversive terms ie. people dying.

A very simple Pavlovian account of this is that the option that is presented as involving sure deaths automatically engages a Pavlovian aversive withdrawal response that is absent for the option involving sure survival (or is maybe turned into an appetitive approach response) that affects its propensity to be chosen. Indeed one can look at the classic trolley moral dilemmas in a similar light (Thomson, 1986). These predictive computations can be quite sophisticated, likely involving model-based as well as model-free systems.

From a neuro-anatomical perspective, framing might be expected to involve neural structures implicated in Pavlovian-instrumental interactions in avoidance and PIT. Indeed, this appears to be the case: we conducted a study involving loss/gain framing of non-risky, alongside risky, financial options, matched for expected value (De Martino et al., 2006).

In the study, participants received a message indicating the amount of money that they would initially receive and then had to choose between a "sure" or a "gamble" option presented in the context of two different "frames". The "sure" option was formulated as either the amount of money retained from the initial starting amount (e.g. keep £20 of a total of £50- "Gain" frame), or as the amount of money lost from the initial amount (e.g. lose £30 of a total of £50- "Loss" frame). The "gamble" option was identical in both frames and represented as a pie-chart depicting the probability of winning or losing. Subjects were risk-averse in the 'Gain' frame, tending to choose the sure option over the gamble option and risk-seeking in the 'Loss' frame, preferring the gamble option (this effect was consistently expressed across different probabilities and starting amounts.

We found that the amygdala correlated with the behavioural influence of the frame on the subjects decisions, being more active when subjects chose in accordance with the frame effect, as opposed to when their decisions ran counter to their general behavioural tendency. Broadly speaking, the data suggest a model in which the framing bias reflects incorporation of a potentially broad range of additional emotional information into the decision process. In evolutionary terms, this mechanism may confer a strong advantage, because such contextual cues may carry useful, if not critical, information. Neglecting such information may ignore the subtle social cues that communicate elements of (possibly unconscious) knowledge that allow optimal decisions to be made in a variety of environments. However, in modern society, which contains many symbolic artefacts and where optimal decision-making often requires skills of abstraction and decontextualization, such mechanisms may be render human choices irrational.

### 9.2.3 Depressive realism.

In comparisons between healthy volunteers and patients with depression, a (not completely uncontroversial) finding is that the volunteers are unduly optimistic about the appetitive value of, and the degree of control they exert over, artificial, experimentally-created environments. By contrast, the depressed subjects make more accurate assessments, and so are more realistic. This phenomenon is called depressive realism (Abramson et al., 1979).

It has been suggested that Pavlovian withdrawal associated with predictions of negative outcomes is an important route to the over-optimism of the volunteers, and that one of the underlying neural malfunctions associated with depression is associated with a weakening of this withdrawal, thereby leading to more accurate, but more pessimistic, evaluations (Huys and Dayan, 2008). Consider a healthy subject entertaining chains of thought about the future. Any chain of thought leading towards a negative outcome engenders a Pavlovian withdrawal response, which may lead to it being terminated or (in the jargon of tree-based search) pruned. Thus if healthy subjects contemplate the future, they will tend to favour samples with more positive outcomes, and will therefore be more optimistic. Given the possibility that this form of Pavlovian withdrawal is mediated by serotonin, as the putative aversive opponent to dopamine (Daw et al., 2002), and the pharmacological suggestion that depressed patients have low effective serotonin levels (Graeff et al., 1996), it is tempting to conclude that this withdrawal mechanism is impaired in the depressed subjects. This would, of course, lead to the basic phenomenon of depressive realism. Indeed, boosting serotonin, which is the ultimate effect of mainline treatment for depression, namely selective serotonin reuptake inhibitors, helps restore the original optimism.

### 9.2.4. Dread.

In an aversive domain, many subjects show an additional sort of impulsivity in the form of dread (Berns et al., 2006). They prefer a larger shock that comes sooner to a weaker shock that comes later, reportedly because of the misery of aversive anticipation. This is exactly the opposite of conventional discounting, which would suggest that more postponed a shock will be, the less it is disliked at the moment. For a more subjective version of this, consider what you would prefer if your dentist discovers a cavity – arranging to have the filling instantly, or booking it for a few weeks' time?

In the study by Berns and colleagues, during the anticipation phase brain regions commonly associated with physical pain are activated, almost as if the anticipation was indeed actually miserable. This idea has been broadened into the more general notion that information (in this case, about a future outcome) can have value itself, a concept that is antithetical to normative Bayesian notions, but is well established in a number of experimental paradigms (Loewenstein, 2006). Subjects behave in odd ways, for instance not collecting free information if it is likely to provide bad news.

Three Pavlovian issues may underlie these facts. First, the activation of the primary pain system is consistent with a Pavlovian phenomenon called stimulus substitution, in which predictors of particular outcomes are treated in many respects like those outcomes themselves. Although the neural foundations of this are not clear, let alone its evolutionary rationale, it is an effect that is widely described, particularly in appetitive circumstances. For instance, the way that a pigeon treats a key which has a Pavlovian association with an appetitive outcome depends directly on whether it is food or water that is predicted. The pecks that result are recognisably associated with the specific outcome itself. Perhaps a model-based form of stimulus substitution leads to an effective overcounting of the temporally distant shock, making the subject prefer the immediate one.

The other two Pavlovian effects are related to those discussed in the context of depressive realism. Not seeking information that is likely to be aversive is exactly akin to not exploring, or actually pruning, paths of thought that are likely to lead to negative outcomes. More subtly, and more speculatively, for the case of dread, maybe the guaranteed prospect of a substantially delayed, future aversive outcome that cannot be controlled has unfortunate model-based and model-free consequences on the Pavlovian mechanism for creating optimism. From a model-based perspective, it creates a prior expectation of environments that are relatively unpleasant because they contain unavoidable aversive outcomes. Such environments are associated with larger average aversive values and so lead to Pavlovian avoidance (Huys and Dayan, 2008). From a model-free perspective, the unavoidable negative outcome might set an adaptation point for the pruning mechanism, and thereby create a circumstance under which substantially more negative paths than normal are explored.

## 9.3 Explicit judgement and value relativity

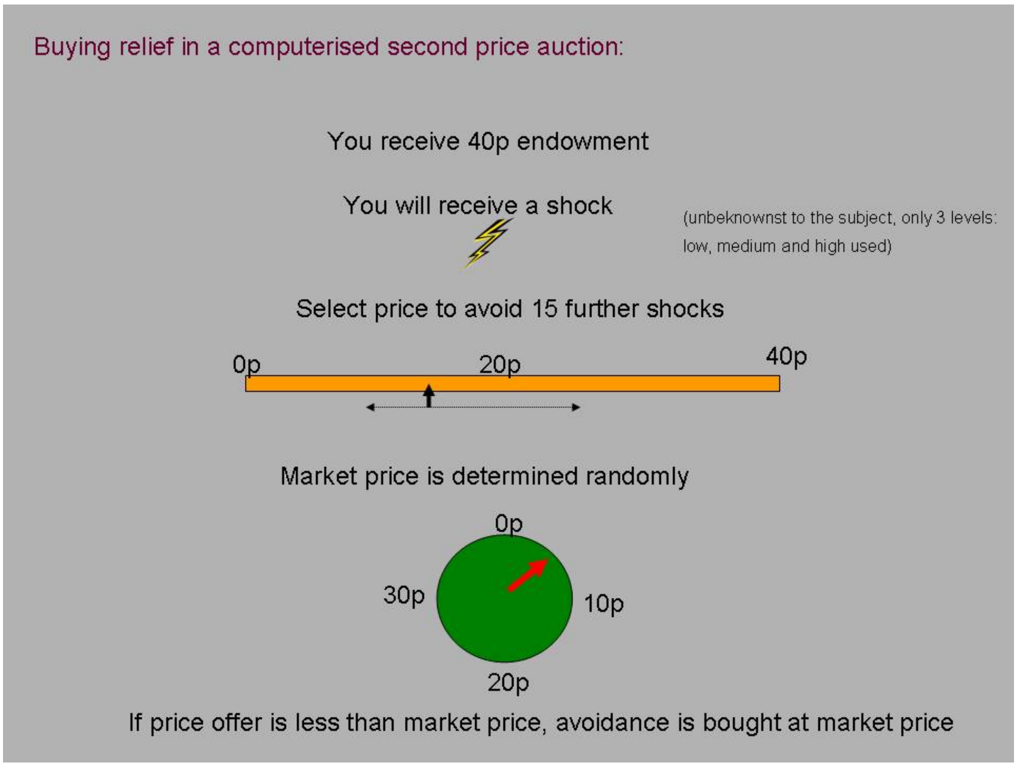
### 9.3.1 Behavioural evidence.

Attaching economic value to aversive states and clinical symptoms (such as pain) is a central issue in political and health economics, and informs issues as diverse as the market price of analgesics, the cost-effectiveness of clinical treatments, compensation for injury, and the response to public hazards. In most cases, the cost of relieving the suffering must be accurately equated with the amount of suffering relieved. Economic theories of valuation generally assume that the prices of such commodities are derived from genuine fundamental values, and that people have robust endogenous preferences and stable trade-offs between goods and money (Shafir & LeBoeuf, 2002). However, the validity of this assumption, and the applicability to health products, is becoming increasingly questioned, and contrasts with an emerging alternative possibility that preferences may be more labile, and predictably so.

This latter view receives support from psychological experiments suggesting that sensory judgments of magnitudes and probabilities are made relative to other recently experienced events, and not bound tightly to some absolute scale. Notably, Ariely, Loewenstein, and Prelec (2003) used annoying sounds, as well as having subjects place their fingers inside a tightening vice, and found that hypothetical willingness-to-pay to avoid prices were typically biased towards price anchors. This resonates with the idea that the mere presence of an option in a choice set may change the way another option is judged; or, more broadly, that preferences are *constructed* afresh in the light of the salient options in each new situation, rather than *revealed* (see Slovic, 1995).

But this conclusion might be premature, because people might not need to know the value of something if they already know its price. Notably, none of the existing studies have tested the preference formation process at its very root—when people experience stimuli or events for very first time and they have to make real monetary valuations, by paying from their own pocket, to obtain or avoid this experience. Indeed, a design with this methodology is a very close approximation to consumer behaviour in the real-world. The stakes are high here, because observing relativistic effects in this context would imply that the price consumers pay (e.g., for health) may be substantially determined by current or recent experiences.

We designed a simple experimental market in which healthy subjects could pay money to avoid an unpleasant (painful) electrical stimulus. In the experiment, subjects received a single electrical stimulus and were then asked to decide how much they were willing to pay, out of their monetary endowment for that trial, in order to avoid fifteen further shocks. If the price offer was more than a randomly determined market price, avoidance was bought at the market price, otherwise the endowment was kept and all shocks had to be endured. This design was analogous to buying pain relief in a computerised ‘second price’ (Becker DeGroot Marshack) auction, and has the characteristic that the only rational bid is according to ones true preferences (all of which was well explained and practiced by each subject).



Buying relief in a computerised second price auction:

You receive 40p endowment

You will receive a shock

(unbeknownst to the subject, only 3 levels: low, medium and high used)

Select price to avoid 15 further shocks

Market price is determined randomly

If price offer is less than market price, avoidance is bought at market price

**Figure 9.1.** Experimental pain-auction task to explore price relativity in an pain 'market', implemented as a Becker DeGroot (second price) auction.

There were 60 such trials and we varied both the shock intensity and the monetary endowment. Unbeknownst to the subject, only 3 pain levels were used: low, medium and high. Furthermore, the levels were grouped into blocks (of 10), such that the different levels co-existed in pairs (Low-Medium, Medium-High, Low-High) (see below).

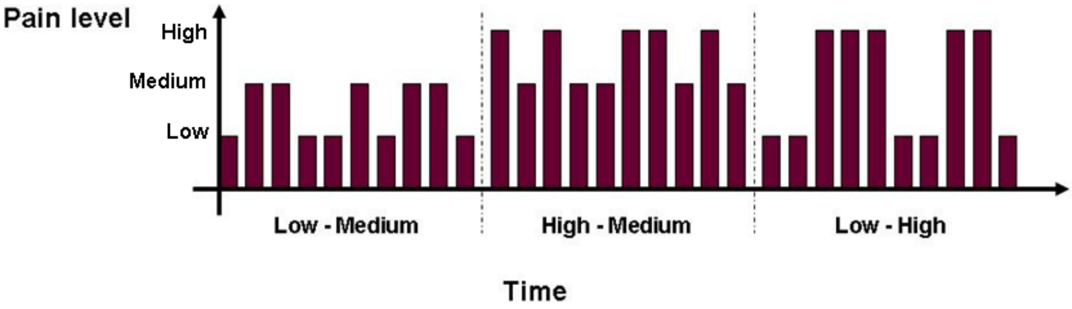


**Figure 9.2.** Experimental pain-auction task: distribution of pain levels was used to create magnitude context effects within blocks

Furthermore, in one group the endowment for each trial was £0.40p, and for second group of subjects it was £0.80p, with subjects randomly assigned to either group.

We observed higher price offers for medium pain relief when experienced in a sequence of trials in which there were many low pain trials (Low-Medium block), compared to when the same pain was experienced in a sequence in which there were many high pain trials (Medium-High block). That is, subjects were willing to pay more for pain relief when it was relatively high compared to the recent average, compared to when it was relatively low, despite the fact the actual magnitude was identical. Furthermore, we observed a striking rescaling of valuation dependent on endowment (40p vs. 80p): higher offers were given when high endowment was received and vice versa.

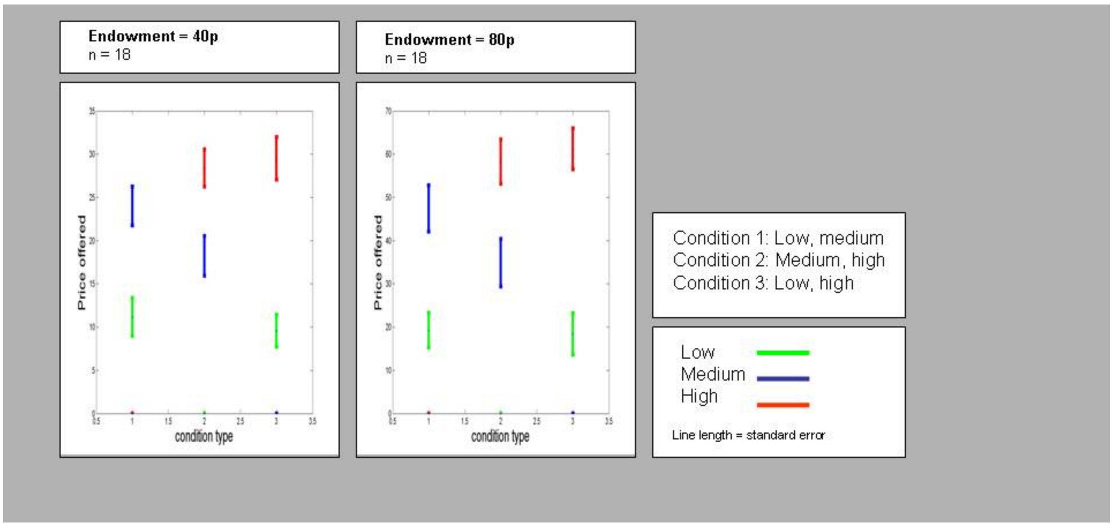


**Figure 9.3.** Experimental pain-auction results: this illustrates context relativity effects induced by both pain level, and cash endowment.

The differential results from subjects with higher cash endowments suggest that people shift (expand) the experimental budget constraint such that they spend roughly a constant fraction of their experimental income on pain relief. A possible argument against the effect of wealth (cash endowment per trial) could be that when people get richer their demand for health might increases, which would explain why people with greater endowments spend more on pain relief. However, this behaviour is a failure of aggregation of the experimental conditions to the rest of the person's health and finances. And if health is special (i.e., not a standard sort of good like chocolate bars for example), it is still strange that the context effect works at the level of the budget for the experiment, because people should be able to integrate the experimental income into their total wealth, and this should not differ between subject groups.

In summary, we found that assessment of pain and demand for pain relief are almost completely relative to a) the experience of pain in the recent past, and b) the current cash-in-hand. Participants were willing to pay a fraction of the 'range' given, regardless of whether the sums they are paying differ by a factor of up to two. What makes pain stimuli especially interesting in this case is the possibility is that people do not know their market price; and it is the knowledge of the market price that determines our willingness to pay for, e.g., a cup of coffee. Once the price is taken away, perhaps we are somewhat lost in our valuations. As a result, economic theories of valuation should not assume that prices of such commodities are derived from genuine fundamental values.

This does not necessarily mean that the brain is inherently poor at forming affective judgements of pain, but it does suggest that our ability to explicitly generate reliable valuations may be sub-optimal, insofar as it is strongly susceptible to contextual effects. With reference to the different value systems discussed above, the necessity to communicate values through an explicit system forces the individual to use a goal-directed, cognitive representation of value. Thus, whereas the judgement lability we see may be less important for the sort of basic decisions we make when interacting with our environment (eg learning not to put our hand in a fire), it may be more problematic when such explicit valuations are the currency by which decisions are made.

Unfortunately, explicit judgments are necessarily required in certain situations. This is the case, firstly, when we are forced to make abstract comparisons between experienced or imagined primary affective states and secondary rewarding ones (such as money). Furthermore, the difficulty in equating such diverse quantities to control purchasing behaviour is confounded by the fact that health products are naturally inhibitory, in that one pays to avoid a certain aversive symptom, rather than pays to receive a positive good. That the product has the positively valenced property of relief has parallels with the nature of the avoidance studied in animal learning theory, in which states that are associated with omission or termination of otherwise aversive events acquire, through inhibitory processes, rewarding valence (add.ref 2-6 )But whereas increasing experience might mitigate this in some situations, it can do no such thing for products which buy relief of never-experienced symptoms, which are a growing commodity in modern preventative healthcare.

Secondly, explicit judgements are required when economists and policy makers need to explicitly quantify adverse clinical states, to make decisions regarding pricing strategy and cost-effectiveness of treatments. Pain is major public health issue, by way its prevalence (around 20% of the general population suffer from clinically significant pain (Eriksen et al., 2003; Macfarlane et al., 2005; NFO World Group, 2007), the cost of analgesics (the global market in analgesics is worth £40 billion), and lost revenue from work absenteeism (in Europe nearly 500 million lost working days every year, costing the economy at least €34 billion). Importantly, pain rarely occurs as an isolated symptom, and usually occurs both in the general symptomatic and temporal context of an illness, provided for instance by the natural course of a disease. Thus, any insights into the structure of human value systems, and its susceptibility to relativistic judgement biases, may have substantial economic consequences when this is taken into account. Future research might usefully explore the stability of valuation for other clinical symptoms, and the effect of other putative contexts such as knowledge and observability of other peoples judgements, which may play an important role in dynamic social markets.

### 9.3.2 Neurobiological insights into value relativity

The above, and other data, leads to two related accounts of how humans generate estimates of the value of goods in transactions.  The first is largely algorithmic, and posits that humans lack stable, long-term representation of the magnitude of value, and judgments are made purely by pair-wise comparisons in an ordinal dimension. This can be formalized by Relative Judgment Models(Laming, 1984;Padoa-Schioppa et al., 2006) and related theories (e.g. the stochastic difference model, multi-alternative decision field theory, adaptation level theory, and range frequency theory (Roe et al., 2001)),and draws support primarily from psychophysical observations. Applying the Relative Judgment Model to value (Stewart et al., 2006), would suggest that initial experience with goods and prices generate the anchors against which subsequent experience is judged.

The second account is computational, and posits that value scales are intact, but that the sensory information from an available option is often inherently uncertain, forcing people have to make inferences (e.g. Bayesian) from all the information presented. Informative and circumstantial cues are thereby exploited for any clues they might harbor regarding the true underlying worth of an option. This view is closely related to theories of perception (Friston, 2003;Kersten and Yuille, 2003), and is well illustrated in vision.

Recent neuroscience research on judgment and decision-making in humans and primates has the capacity to provide evidence of the implementation of these models, and as we show below, evidence exists for both accounts.

#### 9.3.2.1 Relative coding of value

The orbitofrontal cortex has a well-studied role in reward processing, and neuronal activity correlates well with the motivational value of a reward, over-and-above its sensory properties (Padoa-Schioppa and Assad, 2006). For example, activity declines for a reward (or cues that predict a reward) when an individual (human or monkey) is sated with that reward (Critchley and Rolls, 1996;Gottfried et al., 2003), just as it does subjectively. Initial evidence for relative coding came from a classic experiment by Tremblay and Schultz, who presented monkeys with variously preferred juice rewards, and recorded from orbitofrontal neurons while presenting each juice, presented in blocks with one other juice (Tremblay and Schultz, 1999). Critically, neuronal activity depended on whether or not the juice was the preferred in that block, rather than its absolute value (Figure 1). Thus, neurons fired if juice B was presented in blocks in which a less preferred juice (A) was also presented, but not if the alternative was more preferable (juice C). Comparable findings have also been found in human medial orbitofrontal cortex, using an analogous design in an fMRI scanner (Elliott et al., 2008).

A similar pattern occurs with aversive outcomes: if a neutral outcome is presented alongside an electric shock, orbitofrontal neurons respond to the neutral outcome precisely as they do to juice reward presented alongside the neutral outcome (Hosokawa et al., 2007). That is, in both studies, stimuli activate orbitofrontal neurons only when better than their alternative.

More recent studies have shed light in the time course that prescribes the context that provides relative scales. In the previous studies, options were presented individually, with its paired alternative occurring during an individual block of trials (i.e. one block will contain either juice A or B, and another might contain juice B or C). However, if pairs are presented intermixed (i.e. a trial of juice B and C will appear immediately after a trial of A and B), orbitofrontal neurons code absolute value throughout (Padoa-Schioppa and Assad, 2008). In other words, the relative coding of reward seems to exist only between, and not within, blocks.

#### 9.3.2.2 Adaptive scaling

Recording how much better an outcome is in the context of others is clearly useful, and indeed a fully coded version of this is analogous to the prediction error. But theories of relative judgment also suggest that values should *scale* to match the relevant range of magnitudes. Tobler and colleagues (Tobler et al., 2005b) found that just this property was exhibited in dopamine neurons. They conditioned monkeys to predict varying quantities of fruit juice. When they presented cues that predicted two possible, equiprobable amounts, they showed (as expected) that dopamine cell activity coded the relative value of the outcomes (more precisely, the value prediction error), with larger volumes eliciting phasic activations and smaller volumes resulting in deactivations, independent of absolute magnitude. Critically, however, the difference between the activity associated with the higher and lower magnitudes were essentially constant, despite the fact that the volume ranges were substantially different. Thus, the apparent gain, or sensitivity, adapts to the range of magnitudes expected. That such scaling was not seen to the cues themselves, the order of which was unpredictable, suggests that the cues *set* the scale on each occasion, on a trial-by-trial basis. Scaling in the aversive domain has not been studied, to our knowledge.

### 9.3.2.3. Expectation, inference and placebo effects on value

In relative judgment models, contexts may provide anchors to establish scales in determining the relative positions of an option. However, in expectation and "perceptual" models, they actually provide information that influences the experience of it. Expectation effects are well studied in behavioral, psychophysical and economic studies, in both the appetitive and aversive domain. Studies on the latter, which are slightly more extensive, have shown that placebo effects can be reliably induced by either implicit or explicit suggestions that a painful stimulus is less intense than it actually is (or more intense, as in the 'nocebo' effect). Human neuroimaging studies show that brain areas associated with the perception of unpleasantness, the anterior insula cortex and anterior cingulate cortex, show a pattern of activity that reflects the reduced aversive experience induced by expectation despite no change in the actual stimulus, suggesting that the representation of aversiveness is adapted in the brain (Brown et al., 2008a;Wager et al., 2004).

Placebo effects also exist for rewards. De Araujo and colleagues (de Araujo et al., 2005) gave subjects isovaleric acid (which has a cheese-like odor) to subjects in an fMRI scanner, and accompanied it with the words 'cheddar cheese' or 'body odor', exploiting the disconcerting similarity between the two. They found that not only did subjects greatly prefer the scent when labelled ‘cheddar cheese’, but that activity in medial orbitofrontal cortex and rostral anterior cingulate cortex coded this subjective experience. Presumably had they been given the option, they would have paid more money to receive the cheddar cheese smell (or paid to avoid the smelly socks).

Not only can direct suggestions of quality influence subjective experience, but so can prices. Shiv and Ariely and their colleagues studied how the efficacies of products, either energy drinks or an over-the-counter analgesics, yield their behavioral effects depending on their apparent price (Shiv et al., 2005;Waber et al., 2008). They found that energy drinks helped sustain concentration, and analgesics relieved pain more, if they were thought to be more expensive, despite the fact that both products were in fact placebos. This is consonant with the observation that purely sensory judgments are to some extent uncertain, and that subjects use cues (in this case prices) to improve inference.

Recently, the neurobiological basis of this effect has been studied in people. Plassmann and colleagues gave subjects several wines, and provided them with information regarding the retail price of each (Plassmann et al., 2008). Subjects tasting wine they believed to be expensive found it significantly more pleasant than the same wine labelled as being cheap. Neural responses in medial orbitofrontal cortex correlated with the experienced pleasantness, rather than the identity of the wine.

Taken together, these studies show that not only does the subjective experience of a product depend strongly on cues and contexts, be they relevant or irrelevant, but so too does the basic representation of reward or aversive value in the brain.

#### 9.3.2.4 Equating value in transactions

Transactions of any sort involve establishing whether the value of obtaining something compares favorably with the value of losing something else. Since firing rates may not be negative and decreases from baseline firing offer limited resolution, losses and gains may be best encoded by separate populations of neurons. Indeed, this as has been shown in both the orbitofrontal cortex and striatum (Berridge, 2009;O'Doherty et al., 2001;Seymour et al., 2007a).

It remains largely unknown how the brain integrates and compares gain and loss information for explicit values. Knutson and colleagues (Knutson et al., 2007b) have shown that when an explicit trade-off is made between a stated price and an every-day good, there appear to be separate representations of the value of the item to be gained (in nucleus accumbens), and lost (in insula cortex). This leaves open the question of how the trade-off is made. Plassmann and colleagues have shown that subjects' willingness to pay for goods correlates with orbitofrontal cortical activity, consistent with the equation of a common currency of value in this area (since the amount offered will be *lost*) (Plassmann et al., 2007). The fact that the brain area (i.e. the medial orbitofrontal cortex) involved in willingness-to-pay broadly co-localizes with that involved in placebo effects on value, and in the establishment of context-related scales, reaffirms the challenge in understanding exactly how setting up such currency trade-offs proceeds.

The artifacts of the comparison process may be quite striking. That scaling occurs in some form of another is not surprising, and it would be remarkable if neurons encoded accurately the value of goods such as a lunchtime sandwich and the price of our new house on the same scale. If they do indeed adapt suggests, then comparisons across scales might be hazardous. This could offer insight into a classic experiment described by Tversky and Kahneman (Tversky and Kahneman, 1981a), who asked people whether they would spend 20mins to cross town to save $5 on a $15 calculator, or on a $125 jacket. Subjects were far less inclined to do so for the jacket than the calculator, which is clearly absurd, since the absolute amount saved is identical. Clearly, the benefit of adaptive scaling weighs heavily against the inability to integrate across transactions in separate contexts in an individual's daily life.

### 9.3.2.5 Discussion

Neurobiological studies are beginning to provide key insights into why the values people ascribe to goods, and the price they are prepared to pay for them, is often so susceptible to manipulation. First, in given contexts, the brain sets relative scales against which the ordinal position of goods is set. Second, the brain uses available and additional information to help refine judgments of value. Thus, object or price anchors can act in two distinct ways to influence trade decisions. First, they can establish the boundaries and sensitivity (or gain) of a value scale, such that a given transaction will appear *relatively* good or bad. Second, they can appear to provide information about the true worth of a product, and lead the individual to change the judgment and experience of a product.

However, many questions are left open. First, it remains unclear whether absolute value judgments may exist somewhere in the brain. That relative judgements of value are found to exist is not in itself a strong argument that it represents a fundamental characteristic of value encoding, since many related functions, in particular choice, might reasonably be predominantly concerned by how much better or worse one option is to another. Indeed, the striatum has an important role in guiding choice, and hence relative coding and adaptive scaling seen here might occur downstream of *absolute* value coding elsewhere. However, that relative coding is seen in orbitofrontal cortex is more important since this region has a well understood role in basic value coding, although it will be important for future studies to establish whether scaling, in addition, is also a feature of neuronal activity.

Second, evidence that hedonic perception is subject to perceptual priors does not necessarily imply that these influence subsequent decisions (transactions). One of the key insights from behavioral neuroscience to economics has been the realization that there are many interacting value systems that determine behavior [Dayan 2008]. This raises important questions, and limits the generality of conclusions about the findings from existing experiments. Notably, dopaminergic responses are thought to be central to cached Pavlovian and habit like actions, but appear to be less involved in more cognitive, ‘goal-directed’ action (Daw et al., 2005;McClure et al., 2004).

Third, despite good evidence that point predictions provided by cues can seemingly act as inferential priors in hedonic perception, the effect of referential anchors on value within the same modality remains unclear. That is, if you taste a medium quality wine, does this make a subsequently tasted wine taste better or worse? According to a simple Bayesian account, if there is temporal correlation between values, previous stimuli should act as relative attractors. In the absence of this, however, they might be expected to act as repellents, as sometimes seen in adaption effects in other modalities, for instance in colour constancy and tilt illusions (Schwartz et al., 2009). Beyond this, priors might operate at a higher level if, for instance, the brain actually learns *distributions* over values, and uses individual events to learn the parameters of these distributions.

Independently of this, a more straightforward prediction of Bayesian accounts is that certainty or confidence should control the magnitude of expectancy effects. In the appetitive domain, there is some behavioral data indicating that the strength of influence of prior knowledge depends on the amount of experience (Robinson et al., 2007), but the neural basis of this effect has not been established. Recent data from the aversive domain does suggest that greater confidence in prior expectancies results in a greater impact on perception, an effect correlated at a neural level with aversive representations in anterior insula (Brown et al., 2008b). Whether confidence controls placebo effects in markets, both behaviourally or neurally, remains to be tested.

In summary, the way that the brain processes value-related information leaves it vulnerable in many modern day situations. While this is good news for marketing consultants, inspiring various inventive marketing tricks, it is bad news for economists schooled in traditional notion that willingness to pay for goods reflects the inherent, known, and stable values that people ascribe to them.

## 9.4. Aversive motivation in social environments.

### 9.4.1 Introduction

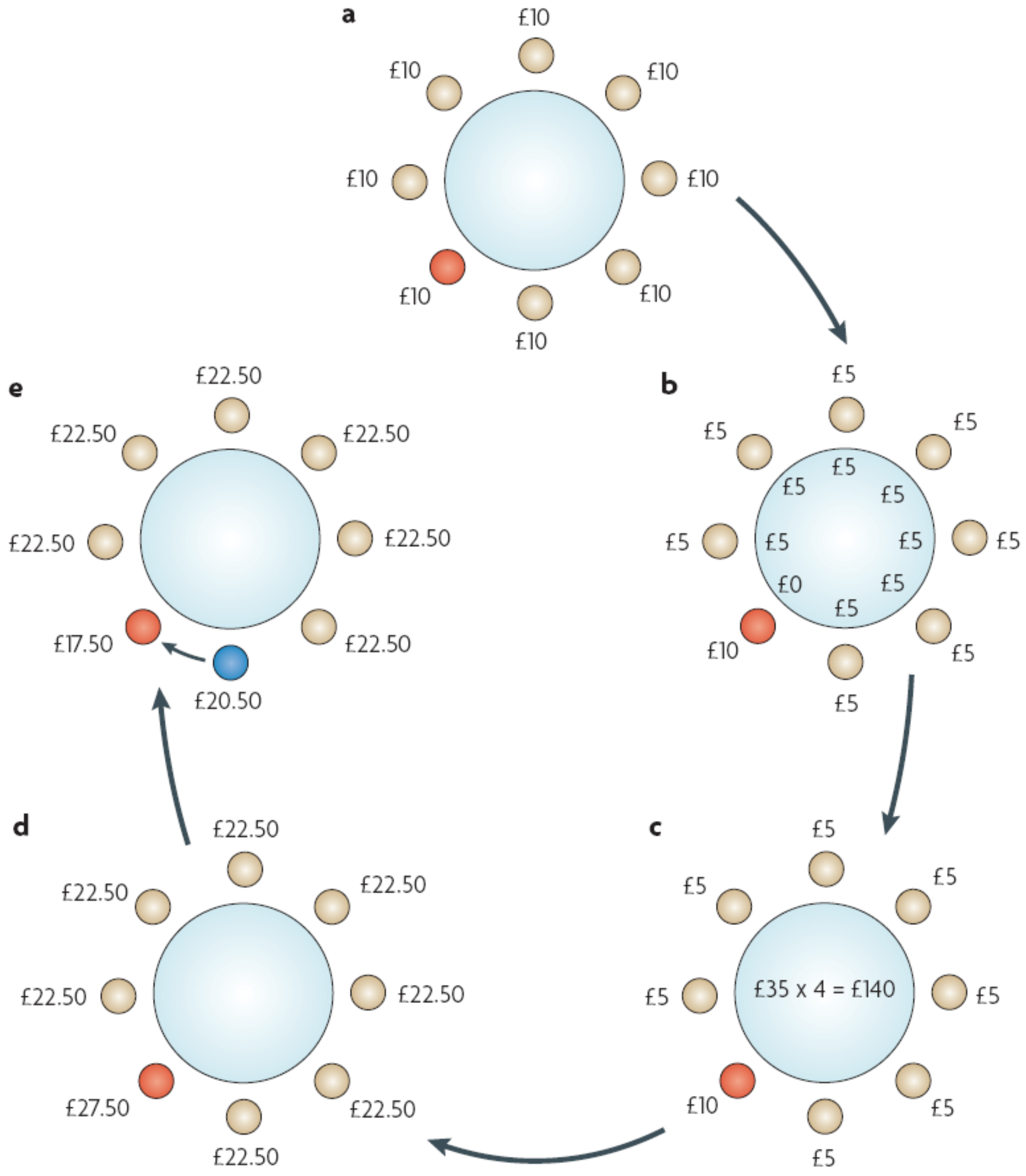
Many social interactions are self-beneficial if we behave positively and pro-cooperatively towards others. Opportunities to benefit from cooperation are widespread, and reflect the extrinsic fact that the natural environment is often best harvested, insofar as rewards can be accrued and threats avoided, by working together. But the decision to cooperate is not always straightforward, as in some situations it leaves us vulnerable to exploitation by others.

Game theory specifies a set of potential social interactions in which outcomes of cooperation and defection systematically differ, allowing both experimentalists and theoreticians to probe an individual's propensity for cooperation in different situations (Camerer, 2003). These outcomes typically vary in the extent to which competitive actions may seem preferable and where a short-sighted temptation to exploit the cooperativeness of others has a capacity to subvert cooperation later. Fortunately, the ability to look beyond the immediate returns of defection towards longer-term cooperation allows humans to escape from otherwise competitive equilibria, and this can be viewed as a hallmark of rational, sophisticated behaviour.

However, humans appear to behave positively towards each other in situations in which there is no capacity to benefit from long-term cooperation: for instance, when they play single games in which they never meet the same opponent again, and when their identities are kept anonymous (Berg et al., 1995;Fehr et al., 1993;Fehr and Fischbacher, 2003). This removes the capacity for both direct reciprocity (tit-for-tat) (Axelrod, 1984;Trivers, 1971), and the ability to earn a cooperative and trustworthy reputation that can be communicated by a third party (Ariely and Norton, 2007a;Bateson et al., 2006;Harbaugh, 1998). Furthermore, they will do this even if it is costly to themselves (Henrich et al., 2006;Xiao and Houser, 2005). From an economic perspective this appears to be genuinely altruistic, being strictly irrational since it incurs a direct personal cost with no conceivable long-term benefit.

Humans also behave *negatively* towards each other in situations in which there is no capacity to benefit, ie they engage in actions that punish others. How punishment might operate in social and reciprocal interactions is illustrated by the free-rider problem. Consider a game in which individual players invest a certain amount of their own money into a central pot (figure 7.4, step 1,2), which is then multiplied by a fixed amount (step 3), and the total amount subsequently divided equally amongst all players (step 4), which they add to the money they *didn't* invest initially. This type of game, termed a public goods game, is similar to many real-life situations, such as a business in which the earnings of each employee depend of the overall turnover of the business. The contribution of each person increases the public good and is beneficial for everyone. More specifically, the overall benefit of the group is bigger than the individual cost of contributing, but this in turn is higher than the direct benefit for the individual. Thus, each individual has also a strong temptation not to contribute in step 2, that is, to free-ride (defect) on the contributions of the rest of the group (see red player) because each individual also profits from the common good, even if he/she does not contribute. If everyone defects, however, cooperation breaks down and the common good is no longer realised. This problem is referred to as the first-order free-rider problem.

**Figure 9.4 The Public Good game**. (see below) Public goods games provide a experimental illustration of the utility of punishment in social economic interactions. In this example, each player receives an initial endowment of £10 (step 1), and contributes a certain proportion toward the public good (step 2), temporarily leaving each with £5. However, the red player - a free rider, contributes nothing, and so remains with £10. The collective contribution is multiplied by a certain amount (4 times in this example), which reflects the overall economic benefit of cooperation (step 3). This amount is then equally divided amongst all players, including the free-rider, who as a result ends up with the most money: £27.5 as opposed to £22.5 (step 4). However, another player (in blue) punishes the free-rider, at personal cost (step 5). Even though this seems irrational in the short term, since it removes the incentive to free-ride in the red-player, the blue player may benefit from future interactions in which the red player cooperates. Thus, in the long run, short term punishment results in long term gain, and reflects a selfish form of reciprocity with repeated interactions. If the blue player does not interact again, however, then punishment becomes altruistic.



Punishment provides a possible solution: if contributing employees start punishing free-riders by fining them (but at personal cost, step 5), the level of cooperation increases again because free-riders want to avoid the cost of being punished (Yamagishi and Sato, 1986). If the punisher knows he/she will interact with the free-rider again, he/she will subsequently benefit from the increased cooperation, and punishment in this case can be viewed as a (long-term) selfish form of reciprocity. However, if the punisher knows that they will not interact with the free-rider again, he/she pays the cost of punishing while others benefit from the free-riders switch to cooperation, and thus punishing becomes altruistic. In reality, as we discuss below, humans punish both selfishly *and* altruistically (Fehr and Gachter, 2002;Yamagishi, 1986).

But a new problem arises: why should individuals endure the costs of punishing free-riders instead of simply cooperating and avoiding the costs of being punished by others? This is the second-order free-rider problem. One solution is to introduce higher levels of punishment, and punish those who do not punish. Boyd and colleagues have proposed another solution, suggesting that human societies maintain punishment by group selection and cultural acquisition and transmission of conformity(Boyd et al., 2003;Boyd and Richerson, 1988;Gintis, 2000). Accordingly, groups with altruistic punishers are able to enforce cooperation norms. With increasing number of punishers the number of defectors in these societies is minimized, as is the cost of punishment. In terms of the ***ultimate*** basis of human reciprocity and cooperation, group selection should favour cooperative groups, allowing punishment and cooperation to evolve. This casts the spotlight upon experimental studies which probe the existence and nature of punishment in both animals and humans.

Arguments against altruistic interpretations of experimentally observed behaviour include suggestions that individuals do not understand the rules of the game, are prone to misbelieve they (or their kin) will interact with opponents again in the future, or falsely infer they are being secretly observed and accordingly act to preserve their reputation in the eyes of experimenters (Smith, 1976). However, the widespread observation of altruism (both rewarding and punishing) across cultures (Henrich et al., 2001a), and within meticulously designed experiments conducted by behavioural economists provide compelling support for its presence as a clear behavioural disposition. Furthermore, in fMRI experiments, altruistic actions correlate with brain activity, suggesting that they derive from some sort of intended or motivated behaviour and are not an expression of mere ‘effector noise’ (ie. decision error)(de Quervain et al., 2004b).

The very existence of altruism raises the difficult question as to why evolution has allowed otherwise highly sophisticated brains to behave so selflessly. This directs attention towards the decision-making systems that subserve economic and social behaviour (Behrens et al., 2009;Lee, 2006;Lee, 2008), and questions whether they are structured in such a way that yields altruism either inadvertently, or necessarily. The broader consequence is that if they do, then this reframes the question regarding the ultimate (evolutionary) causes of altruism towards the evolution of these very decision systems, and away from the phenomenological reality of altruism per se.

### 9.4.2 Experimental observations of punishment in animals and humans.

Animals not infrequently behave negatively to one another. In many cases, this is driven by an immediate selfish benefit to the animal (or its kin) effecting the behaviour — for example, assertion of dominance, the establishment of mating bonds, theft, parental–offspring conflicts and retaliation (Clutton-Brock and Parker, 1995). In some situations, food-sharing is increased by harassment, although whether this represents cooperation is unclear (Stevens and Hauser, 2004). For example, the sharing rate in chimpanzees and squirrel monkeys increase with increasing acts of harassment(Stevens, 2004). However, punishment is observed in some situations where it seems more likely to preserve or promote cooperation. For instance, chimpanzees attack allies that do not support them in third party conflicts (De Waal, 1998), and queen naked mole rats will attack workers whom they judge lazy (Reeve, 1992). Cases such as these highlight behaviour that influences future, non-immediate actions of others, rather than conferring immediate self-benefit. These dispositions might represent the evolutionary precursor of more complex and ultimately altruistic punitive behaviours widely seen in humans (Stevens, 2004).

In addition to more simple (defensive and retaliative) forms of punishments, humans also clearly use punishment to motivate others to cooperate (Shinada et al., 2004). One of the classic experimental demonstrations was provided by Yamagishi, who studied cooperation in a public goods game (Yamagishi, 1986).

He showed that sanctioning by means of financial penalties increased cooperation in subsequent rounds of the game, and in comparison to games in which there was no opportunity for punishment.

The existence of *altruistic* punishment as a proximate intentional motivation in humans is evident by demonstrations that people are willing to incur a personal cost solely to punish others whom they consider to have behaved unfairly. The simplest illustration occurs in the Ultimatum game, where a player decides whether to accept a proposed split, offered by another player, of a central pot of money. Typically, unequal (<20%) splits are rejected, which cause both proposer and responder to leave empty handed. This institution of costly, altruistic punishment for unfair behaviour seems to be ubiquitous across widely different societies and cultures (Henrich et al., 2001b;Henrich, 2006).

Altruistic punishment robustly promotes cooperation (Boyd and Richerson, 1992;Fehr and Gachter, 2000;Fehr and Gachter, 2002). For example, Gürerk, Irlenbusch and Rockenbach allowed subjects to choose between playing public goods games in institutions (societies) which did or did not offer the opportunity to punish and reward others (Gurerk et al., 2006). Even though subjects initially tended towards those institutions where they couldn't be punished, the pay-offs in these groups declined as they became dominated by free-riders, and most subjects switched to play in sanctioning games where the overall level of cooperation progressively increased. Subsequent studies have indicated that cooperation may be even more robust if altruistic punishment is combined with altruistic reward, in which cooperativeness of others is rewarded (at personal cost) (Andreoni et al., 2003).

The proposed importance of cultural norms in driving behaviour predict that individuals ought to be motivated to reward and punish those who adhere to or transgress norms towards others, even when they themselves are not involved (Bendor and Swistak, 2001). These situations are captured by third-party punishment games, in which an observer witnesses the interactions of two other players. For example, Fehr and Fischbacher implemented a third-party punishment game in the context of simultaneous prisoner's dilemma task (Fehr and Fischbacher, 2004b): a subject observed the behaviour of two players during the game, and was subsequently given the option to punish at personal cost. Players who cooperated were almost never punished, whereas almost 50% of subjects punished players who defected when their partner cooperated. When *both* players defected, the punishment rate decreased to 21%. This asymmetry appears to reflect the norm of conditional cooperation, which prescribes that cooperation is assumed if the other player cooperates, whereas defection is considered a more legitimate (less unfair) response in the face of defection by others. Accordingly, unilateral defection is sanctioned more strongly than mutual defection (Fehr and Fischbacher, 2004a). Once a group establishes a strong reciprocating culture, interaction with other forms of (selfish) reciprocity may mean that the costs of altruistically punishing become relatively small (Boyd et al., 2003;Rockenbach and Milinski, 2006). In effect, the *threat* of punishment may become effective in maintaining cooperation.

### 9.4.3 Neuroimaging studies in humans.

Recently, fMRI has been used to probe the neurobiological correlates of human cooperative behaviour in game theoretic experiments. In particular, several studies have addressed the neurobiological correlates of fairness and punishment, establishing findings which begin to shed light onto the underlying basis of punishing actions. Sanfey and colleagues studied the response to fair and unfair offers in an ultimatum game (Sanfey et al., 2003). They found that activity in the anterior insula correlated with the receipt of an unfair offer, which was greater when playing a human as opposed to a computerised opponent, and was greater still with increasingly unfair offers. Impressively, this activity predicted subjects subsequent decisions to reject the offer, effectively (altruistically) punishing their opponent. This study also identified activity in dorsolateral prefrontal cortex (DLPFC) in relation to fair offers, but not correlated with either the degree of unfairness, hinting that it might adopt a more modulatory role. This proposition was supported by a study from Knoch and colleagues, who disrupted DLPFC activity with transcranial magnetic stimulation (TMS) during the Ultimatum game (Knoch et al., 2006). They found that TMS applied to the right, but not left, DLPFC reduced subjects’ decisions to reject unfair offers. This behaviour was specific to human opponents, insensitive to the magnitude of the offer, and independent of subjective verbal ratings of unfairness.

These findings lead to the question of how the representation of the aversive motivational value of unfairness is linked to behavioural decisions to punish. Ultimately, the individual must choose between two outcomes: the financial value of accepting the offer, and the retributive value of punishing the opponent. We designed a task aimed to identify brain areas associated with retributive value, by looking at the response to cues that predicted that opponents would receive painful electric shocks (Singer et al., 2006). We compared brain activity elicited when the cues signalled that a fair, or unfair, opponent would receive either a high or low intensity shock, where the degree of fairness was associated with previous play in a sequential prisoners dilemma game. Medial orbitofrontal cortex and nucleus accumbens were activated when cues indicated imminent high intensity shock to unfair players, and this activity correlated with subjects subjective feelings of anger and retribution. These findings, which were accompanied by compensatory decreases in empathic neural responses, highlight the flexible representation of retributive goals in orbitofrontal cortex, similar to that seen for primary rewards.

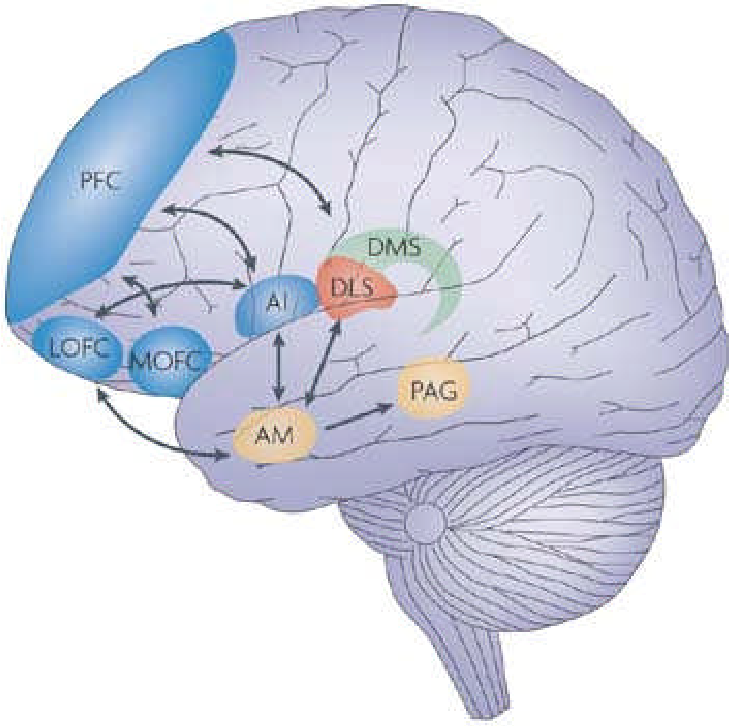
While passive tasks such as this are adequate for identifying brain areas associated with retributive motivational states, they offer little insight into the question of control: that is, which brain areas are involved in learning and executing actions to bring about punishment? De Quervain and colleagues gave subjects the opportunity to punish unfair opponents, at personal cost, in an anonymous trust game(de Quervain et al., 2004a). Using positron emission tomography (PET), they found that activity in dorsal striatum was associated with altruistic punishment acts, with greater activation associated with more severe punishments (which were tied to greater personal losses).

### 9.4.4. A neurobiological model.

Taken together with an understanding of the basic motivation and action selection, these findings allow one to sketch a neurobiological model of punishment. In the simplest case, if an aversive outcome appears to be directly and predictively associated with another individual, it would seem likely to invoke a Pavlovian mechanism, centred on the amygdala, that may present a relatively pre-potent or impulsive route to punishment. This pathway may direct retaliative responses towards that individual, mediated in part via aggression related areas such as the periaqueductal grey. Furthermore, this amygdala-dependent pathway may have a central role in guiding escape and avoidance from future interactions with that individual, contributing to subsequent ostracism.

The amygdala may exploit functional connectivity with the lateral orbitofrontal cortex and anterior insula, which may be necessary for more sophisticated, context dependent aversive representations, for instance those relating to fairness. In principle, one can import fairness-related outcomes onto Dickinson and Dearing's 'Konorskian' model (Figure 1) to specify the full range of excitatory–inhibitory fairness-related outcomes (and predictors) (Figure 4). This would predict that the anterior insula is similarly involved in representing retributive inhibitors – that is, outcomes and predictive cues associated with the frustration of seeing a free-rider unpunished. However, at the current time we know relatively little about how the brain represents observed norms of cooperative behaviour in a way that allows judgement of the fairness of others' behaviour (Fehr and Fischbacher, 2004a;Moll et al., 2005).

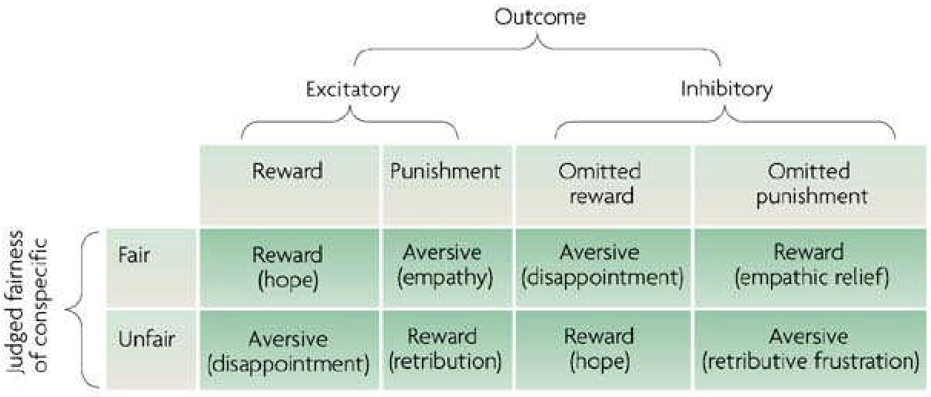
Beyond these simple aversive responses, instrumental control may be dependent on an appropriate representation of the appetitive retributive value of outcomes associated with successful punishment, represented in the medial orbitofrontal cortex. This appetitive value may reinforce punishing actions (or avoidance actions), through reciprocal connections with dorsal striatum, in a similar manner to primary rewards. Furthermore, reinforcement may arise from complex models of future reciprocal interactions involving more widespread areas or prefrontal cortex: this may include theory of mind areas (anterior paracingulate cortex, the superior temporal sulci and the temporal poles) likely to be involved in representing the policies of others (Brunet et al., 2000;Gallagher et al., 2002;Gallagher and Frith, 2003), anterior cingulate cortical subregions involved in representing agency (Tomlin et al., 2006), and more anterior prefrontal cortical areas involved in model-building and resolution of partial observability (Yoshida and Ishii, 2006). Ultimately, in repetitively predictable situations, such actions may become habitual responses to unfairness.



Nature Reviews | Neuroscience

**Figure 9.5.** A neurobiological tri-partite model of social punishment. Impulsive, predominantly Pavlovian punishment may centre on an amygdala-based circuit (depicted in yellow), in which there is associative learning between other individuals (which act as cues) and aversive outcomes. Aversive outcomes may input directly to the amygdala (for example, from brainstem nuclei associated with primitive aversive representations, such as pain[29]), or through more complex aversive representations in the anterior insula (AI) and lateral orbitofrontal cortex (LOFC). This pathway might also be important for avoidance and ostracism. Instrumental punishment may involve striatal-mediated reinforcement of actions that lead to appetitive retributive goals. This appetitive representation (depicted in blue) may involve the medial orbitofrontal cortex (MOFC), and might result from forward-planning of future interactions in broader areas of the prefrontal cortex (PFC) involved in theory or mind, agency, hidden state-estimation and working memory. Goal-directed actions may reinforce action through the dorsomedial striatum (DMS, green). Habit-based actions might reinforce action through dorsolateral striatum (DLS, red), possibly utilizing a

dopamine-dependent circuit via the substantia nigra and ventral tegmental area. PAG, periaqueductal grey.


Outcome
Excitatory
Inhibitory
Judged fairness of conspecific

| | Reward | Punishment | Omitted reward | Omitted punishment |
|---|---|---|---|---|
| Fair | Reward (hope) | Aversive (empathy) | Aversive (disappointment) | Reward (empathic relief) |
| Unfair | Aversive (disappointment) | Reward (retribution) | Reward (hope) | Aversive (retributive frustration) |

Nature Reviews | Neuroscience

Figure 9.6 This figure extends the Dickenson and Dearing's 'Konorskian' motivational model[13] to incorporate social reinforcement made with respect to judgements of fairness. When affective outcomes are observed in conspecifics who are fair (or who are kin), the motivational value is congruent with the observer. If the individual is judged to be unfair, then the pattern of value is reversed. This illustrates the full spectrum of prosocial motives according to predicted or omitted outcomes, or their predictors.

### 9.4.5 Altruistic punishment.

The retributive value of punishment may arise from potentially sophisticated forward modelling of future interactions. But this leads to the question of how *altruistic* goals are acquired, if they, by definition, ultimately result in personal cost. There are several possibilities. First, they may reflect a misassumption that future interactions are not improbable (not unreasonable in smaller societies in human evolutionary history). Second, they could reflect the anticipated prospect that kin, possibly in subsequent generations, will interact with the individual being punished. Third, if punishment from ‘selfish’ reciprocal (goal-orientated) action reliably results in eventual long-term payoffs, more proximal states following punishment may be reinforced both through habit based learning, and through sequential Pavlovian learning (Vlaev and Chater, 2006). This latter process allows the state immediately following punishment to acquire an appetitive value, which may then independently reinforce other actions (through conditioned reinforcement). Both these forms of control will be insensitive to the possibility that in some situations the outcome is altruistic. Fourth, it is possible that learning mechanisms involved when observing others punishing, in situations which may not necessarily be altruistic, generalise across situations in which it is. Given that many selfish reciprocal punishing actions may stem form a long-term view of future interactions, the eventual benefits of an action are likely to be frequently obscure to a naive observer. In other words, the appetitive value of retributive states and actions might be purely imitated or inferred through observation, since the observer does not have access to the eventual goals in the mind of the individual being observed. Thus, the motivation to punish unfair individuals may be acquired across states in a way that *assumes* eventual outcomes. Elsewhere, we detail precisely how such learning mechanisms might yield altruism from both habitization and observation (Seymour et al., 2009). Fifth, and in a similar manner, the value of punishment may be taught by experts to non-experts (for example, from parents to offspring, or from dominant to subordinate individuals). In this case, the appetitive value of punishment may be intricately tied in with cultural concepts of morality and justice.

Thus, the very nature of action systems, both those involved in individual *and* observational learning, may have an inherent tendency to generalise non-altruistic to altruistic actions (Seymour et al., 2009). This suggests that there is no reason to assume that altruistic punishment should necessarily be hard-wired as inherited intrinsic motivational goals (that is, as an unconditioned appetitive stimulus) in the same manner as primary rewards. However, neither does it exclude the possibility. Future research may help resolve both the role of learning and early development in the acquisition of altruistic behaviour.

Clearly, there are many potentially complex ways in which punishing behaviour, including altruistic punishment, might be acquired, and the nature of this acquisition governs the types of action by which it is mediated. Although this says nothing about *why* such behaviour should have evolved (that is, the ultimate basis of different forms of punishment), it illustrates (proximately) *how* they might be based on the operation and, importantly, the interaction of different learning systems. Furthermore, this complexity illustrates the difficulty evolutionary models face. Since underlying learning and decision making processes are not solely concerned with punishment behaviour, such models need to take into account the other behaviours that these systems subserve, many of which are not related to reciprocity and cooperation. This difficulty may be similarly evident in other apparently irrational punishment-related behaviour, such as self-punitive actions and reciprocal aggression. Thus, future models may need to take a more generic approach to understanding the interaction between evolution and learning(Ackley and Littman, 1991).

### 9.4.6 Conclusions

Punishment, in its various forms, is likely to have played a key role in shaping the dynamics of social interaction in many species, and humans in particular. Although many aspects of our neurobiological model are speculative, punishment is likely to involve the integration of a number of distinct representation, learning and action systems. Whatever the neural mechanism, the affirmation that punishment, including altruistic punishment, substantially promotes cooperation in human societies seems firm. Critical to furthering our knowledge will be understanding the behavioural and neurobiological basis of cultural and observational learning, sequential learning, and model-based learning and planning in the context of other agents. This may be crucial to gaining neurobiological insight into how apparently altruistic behaviours are acquired, as well as shedding light onto more complex social aspects of punishment, such as the arbitration, policing, and the role of hierarchies.

## Reference List

1. Abramson,L.Y., Metalsky,G.I., and Alloy,L.B. Judgment of contingency in depressed and nondepressed students: sadder but wiser? J.Exp.Psychol.Gen 108[4], 441-485. 1979.

2. Ackley,D.H. and Littman,M.L. (1991) Interactions between learning and evolution," in *Artificial Life II, SFI Studies in the Sciences of Complexity, Vol. X*, C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, Eds. Reading, MA: Addison-Wesley. 487-509.

3. Adams,D.B. (2006). Brain mechanisms of aggressive behavior: an updated review. Neurosci.Biobehav.Rev. *30*, 304-318.

4. Altier,N. and Stewart,J. (1999). The role of dopamine in the nucleus accumbens in analgesia. Life Sci. *65*, 2269-2287.

5. Amaral,D.G. and Price,J.L. (1984). Amygdalo-Cortical Projections in the Monkey (Macaca-Fascicularis). Journal of Comparative Neurology *230*, 465-496.

6. Andreoni,J., Harbaugh,W., and Vesterlund,L. (2003). The carrot or the stick: Rewards, punishments, and cooperation. American Economic Review *93*, 893-902.

7. **Ariely,D. and Norton,M.I. (2007b). Psychology and experimental economics: A gap in abstraction. Current Directions in Psychological Science *16,* 336-339.**

8. **Ariely,D. and Norton,M.I. (2007a). Psychology and experimental economics: A gap in abstraction. Current Directions in Psychological Science *16,* 336-339.**

9. **Atnip,G.W. (1977). Stimulus and response reinforcer contingencies in autoshaping, operant, classical and omission training procedures in rats. J.Exp.Anal.Behav. 28, 56-69. 1977.**

10. **Axelrod,R. (1984). The Evolution of Cooperation. (**New York: Basic Books).

11. **Azrin,N.H. Some effects of two intermittent schedules of immediate and non-immediate punishment. Journal of Psychology 42, 3-21. 1956.**

12. **Azrin,N.H. (1960). Effects of punishment intensity during variable-interval reinforcement. J.Exp.Anal.Behav. *3,* 123-142.**

13. **Azrin,N.H., Holz,W.C., and Hutchinson,R.R. Fixed-ratio escape reinforcement. Journal of the Experimental Analysis of Behavior 6, 141-148. 1963.**

14. **B.S.Kapp, P.J.Whalen, W.F.Supple, and J.P.Pascoe. (1992). Amygdaloid contributions to conditioned arousal and sensory processing. In: J.P. Aggleton, Editor, The amygdala: neurobiological aspects of emotion, memory, and mental dysfunction, Wiley-Liss, New York, pp. 229-254. 1992.**

15. **B.Setlow, P.C.Holland, and M.Gallagher. (2000). Involvement of a basolateral amygdala complex-nucleus accumbens system in appetitive Pavlovian second-order conditioning. Soc Neurosci Abstr 26, p. 1504.**

16. **Balleine,B. (1992). Instrumental performance following a shift in primary motivation depends on incentive learning. J.Exp.Psychol.Anim Behav.Process *18,* 236-250.**

17. **Balleine,B.W. (2005). Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. Physiol Behav. *86,* 717-730.**

18. **Balleine,B.W. and Dickinson,A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology *37,* 407-419.**

19. **Balleine,B.W., Killcross,A.S., and Dickinson,A. (2003). The effect of lesions of the basolateral amygdala on instrumental conditioning. Journal of Neuroscience *23,* 666-675.**

20. **Balleine,B.W. and Killcross,S. (2006). Parallel incentive processing: an integrated view of amygdala function. Trends in Neurosciences *29,* 272-279.**

21. **Baron,A. (1965). Delayed punishment on a runway response. J.Comp Physiol Psychol. *60,* 131-134.**

22. **Barto AG, Sutton RS, and Anderson CW. (1993). Neuronlike elements that can solve difficult learning problems. IEEE Transactions on Systems, Man, and Cybernetics 13, 834-846.**

23. **Barto,A.G. Adaptive critic and the basal ganglia. (1995). In JC Houk, JL Davis & DG Beiser (Eds). *Models of information processing in the basal ganglia* (pp. 215-232). Cambridge: MIT press.**

24. **Barto,A.G., Sutton,R.S., and Watkins,C.J.C.H. (1990). Learning and sequential decision making. In M Gabriel & J Moor, eds. *Learning and Computational Neuroscience: Foundations of Adaptive Networks.* Cambridge, MA: MIT press. 539-602.**

25. **Bateson,M., Nettle,D., and Roberts,G. (2006). Cues of being watched enhance cooperation in a real-world setting. Biology Letters *2,* 412-414.**

26. **Baxter,M.G. and Browning,P.G.F. (2007). Two wrongs make a right: Deficits in reversal learning after orbitofrontal damage are improved by amygdala ablation. Neuron *54,* 1-3.**

27. **Baxter,M.G. and Murray,E.A. (2002). The amygdala and reward. Nat.Rev.Neurosci. *3,* 563-573.**

28. **Baxter,M.G., Parker,A., Lindner,C.C.C., Izquierdo,A.D., and Murray,E.A. (2000). Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. Journal of Neuroscience *20,* 4311-4319.**

29. **Bayer,H.M. and Glimcher,P.W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron *47,* 129-141.**

30. **Becerra,L., Breiter,H.C., Wise,R., Gonzalez,R.G., and Borsook,D. (2001). Reward circuitry activation by noxious thermal stimuli. Neuron *32,* 927-946.**

31. **Bechara,A., Damasio,H., Damasio,A.R., and Lee,G.P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. J.Neurosci. *19,* 5473-5481.**

32. **Bechara,A., Tranel,D., and Damasio,H. (2000). Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. Brain *123,* 2189-2202.**

33. **Behrens,T.E., Hunt,L.T., and Rushworth,M.F. (2009). The computation of social behavior. Science *324,* 1160-1164.**

34. **Bellman,R. *Dynamic Programming.* Princeton, NJ:Princeton University Press. 1957.**

35. **Belova,M.A., Paton,J.J., Morrison,S.E., and Salzman,C.D. (2007). Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. Neuron *55,* 970-984.**

36. **Bendor,J. and Swistak,P. (2001). The evolution of norms. American Journal of Sociology *106,* 1493-1545.**

37. **Berg,J.E., Dickhaut,J., and McCabe,K. (1995). Trust, reciprocity, and social history. Games and Economic Behavior *10,* 122-142.**

38. **Berns,G.S., Chappelow,J., Cekic,M., Zink,C.F., Pagnoni,G., and Martin-Skurski,M.E. (2006). Neurobiological substrates of dread. Science *312,* 754-758.**

39. **Berridge,K.C. (2009). 'Liking' and 'wanting' food rewards: Brain substrates and roles in eating disorders. Physiology & Behavior *97,* 537-550.**

40. **Bersh,P.J. and Lambert,J.V. (1975). Discriminative Control of Free-Operant Avoidance Despite Exposure to Shock During Stimulus Correlated with Nonreinforcement. Journal of the Experimental Analysis of Behavior *23,* 111-120.**

41. **Bertsekas,D.P. (1995). Dynamic Programming and Optimal Control, Athena Scientific.**

42. **Biederman,G. (1968). Discriminated Avoidance Conditioning - Cs Function During Avoidance Acquisition and Maintenance. Psychonomic Science *10,* 23-&.**

43. **Biederman,G., D'Amato,M.R., and Keller,D. (1964). Facilitation of discriminated avoidance learning by dissociation of CS and manipulandum. Psychonom.Sci. 1, 229-230. 1964.**

44. **Bitsios,P., Szabadi,E., and Bradshaw,C.M. (2004). The fear-inhibited light reflex: importance of the anticipation of an aversive event. Int.J.Psychophysiol. *52,* 87-95.**

45. **Blaisdell,A.P., Sawa,K., Leising,K.J., and Waldmann,M.R. (2006). Causal reasoning in rats. Science *311,* 1020-1022.**

46. **Bolles,R.C., Holtz,R., Dunn,T., and Hill,W. (1980). Comparison of stimulus learning and response learning in a punishment situation. Learn Motiv 11, 78-96.**

47. **Boyd,R., Gintis,H., Bowles,S., and Richerson,P.J. (2003). The evolution of altruistic punishment. Proc.Natl.Acad.Sci.U.S.A *100,* 3531-3535.**

48. **Boyd,R. and Richerson,P.J. (1988). The evolution of reciprocity in sizable groups. J.Theor.Biol. *132,* 337-356.**

49. **Boyd,R. and Richerson,P.J. (1992). Punishment Allows the Evolution of Cooperation (Or Anything Else) in Sizable Groups. Ethology and Sociobiology *13,* 171-195.**

50. **Brandon,S.E., Vogel,E.H., and Wagner,A.R. (2003). Stimulus representation in SOP: I. Theoretical rationalization and some implications. Behav.Processes *62,* 5-25.**

51. **Breiter,H.C., Aharon,I., Kahneman,D., Dale,A., and Shizgal,P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. Neuron *30,* 619-639.**

52. **Brown,C.A., Seymour,B., Boyle,Y., El Deredy,W., and Jones,A.K. (2008a). Modulation of pain ratings by expectation and uncertainty: Behavioral characteristics and anticipatory neural correlates. Pain *135,* 240-250.**

53. **Brown,C.A., Seymour,B., El Deredy,W., and Jones,A.K. (2008b). Confidence in beliefs about pain predicts expectancy effects on pain perception and anticipatory processing in right anterior insula. Pain *139,* 324-332.**

54. **Brown,P. and Molliver,M.E. (2000). Dual serotonin (5-HT) projections to the nucleus accumbens core and shell: Relation of the 5-HT transporter to amphetamine-induced neurotoxicity. Journal of Neuroscience *20,* 1952-1963.**

55. **Brunet,E., Sarfati,Y., Hardy-Bayle,M.C., and Decety,J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. Neuroimage. *11,* 157-166.**

56. **Buchel,C. and Dolan,R.J. (2000). Classical fear conditioning in functional neuroimaging. Curr.Opin.Neurobiol. *10,* 219-223.**

57. **Buchel,C., Dolan,R.J., Armony,J.L., and Friston,K.J. (1999). Amygdala-hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging. J.Neurosci. *19,* 10869-10876.**

58. **Buchel,C., Holmes,A.P., Rees,G., and Friston,K.J. (1998). Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. Neuroimage. *8,* 140-148.**

**59. Bull,J.A. and Overmier,J.B. (1968). Additive and Subtractive Properties of Excitation and Inhibition. Journal of Comparative and Physiological Psychology *66,* 511-&.**

**60. Burns,L.H., Robbins,T.W., and Everitt,B.J. (1993). Differential-Effects of Excitotoxic Lesions of the Basolateral Amygdala, Ventral Subiculum and Medial Prefrontal Cortex on Responding with Conditioned Reinforcement and Locomotor-Activity Potentiated by Intraaccumbens Infusions of D-Amphetamine. Behavioural Brain Research *55,* 167-183.**

**61. Cabanac,M. (1971). Physiological role of pleasure. Science *173,* 1103-1107.**

**62. Cador,M., Robbins,T.W., and Everitt,B.J. (1989). Involvement of the Amygdala in Stimulus Reward Associations - Interaction with the Ventral Striatum. Neuroscience *30,* 77-86.**

**63. Calder,A.J., Lawrence,A.D., and Young,A.W. (2001). Neuropsychology of fear and loathing. Nat.Rev.Neurosci. *2,* 352-363.**

**64. Camerer,C. (1995). Individual decision making, in John H. Kagel, and Alvin E. Roth, ed.: *The Handbook of Experimental Economics* ~Princeton University Press, Princeton NJ.**

**65. Camerer,C., Loewenstein,G., and Prelec,D. (2005). Neuroeconomics: How neuroscience can inform economics. Journal of Economic Literature *43,* 9-64.**

**66. Camerer,C.F. (2003). Behavioural Game Theory: Experiments in Strategic Interaction. Princeton University Press).**

**67. Camp,D.S., Raymond,G.A., and Church,R.M. (1967). Temporal relationship between response and punishment. J.Exp.Psychol. *74,* 114-123.**

**68. Cardinal,R.N., Parkinson,J.A., Hall,J., and Everitt,B.J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci.Biobehav.Rev. *26,* 321-352.**

**69. Cardinal,R.N., Winstanley,C.A., Robbins,T.W., and Everitt,B.J. (2004). Limbic corticostriatal systems and delayed reinforcement. Adolescent Brain Development: Vulnerabilities and Opportunities *1021,* 33-50.**

**70. Carpenter,L.L., Anderson,G.M., Pelton,G.H., Gudin,J.A., Kirwin,P.D., Price,L.H., Heninger,G.R., and McDougle,C.J. (1998). Tryptophan depletion during continuous CSF sampling in healthy human subjects. Neuropsychopharmacology *19,* 26-35.**

71. **Carter,R.M., O'Doherty,J.P., Seymour,B., Koch,C., and Dolan,R.J. (2006). Contingency awareness in human aversive conditioning involves the middle frontal gyrus. Neuroimage. *29,* 1007-1012.**

72. **Chudler,E.H. and Dong,W.K. (1995). The role of the basal ganglia in nociception and pain. Pain *60,* 3-38.**

73. **Church,R.M. (1969a). Response suppression, in Punishment and Aversive Behavior (B. A. Campbell and R. M. Church, eds.), Appleton, New York.**

74. **Church,R.M. (1969b). Response suppression. In: B.A. Campbell and R.M. Church, Editors, *Punishment and aversive behavior,* Appleton-Century-Crofts, New York.**

75. **Church,R.M., Raymond,G.A., and Beauchamp,R.D. (1967). Response suppression as a function of intensity and duration of a punishment. J Comp Physiol Psychol 1, 39-44.**

76. **Clark,R.E. and Squire,L.R. (1998). Classical conditioning and brain systems: the role of awareness. Science *280,* 77-81.**

77. **Clarke,H.F., Dalley,J.W., Crofts,H.S., Robbins,T.W., and Roberts,A.C. (2004). Cognitive inflexibility after prefrontal serotonin depletion. Science *304,* 878-880.**

78. **Clarke,H.F., Walker,S.C., Dalley,J.W., Robbins,T.W., and Roberts,A.C. (2007). Cognitive inflexibility after prefrontal serotonin depletion is behaviorally and neurochemically specific. Cereb.Cortex *17,* 18-27.**

79. **Clutton-Brock,T.H. and Parker,G.A. (1995). Punishment in animal societies. Nature *373,* 209-216.**

80. **Cook,L. and Catania,A.C. (1964). Effects of drugs on avoidance and escape behaviour. Fed.Proc. *23,* 818-835.**

81. **Cools,R., Roberts,A.C., and Robbins,T.W. (2008). Serotoninergic regulation of emotional and behavioural control processes. Trends Cogn Sci. *12,* 31-40.**

82. **Corbit,L.H. and Balleine,B.W. (2005). Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of Pavlovian-instrumental transfer. J.Neurosci. *25,* 962-970.**

83. **Corbit,L.H., Muir,J.L., and Balleine,B.W. (2001). The role of the nucleus accumbens in instrumental conditioning: Evidence of a functional dissociation between accumbens core and shell. J.Neurosci. *21,* 3251-3260.**

**84. Coricelli,G., Critchley,H.D., Joffily,M., O'Doherty,J.P., Sirigu,A., and Dolan,R.J. (2005). Regret and its avoidance: a neuroimaging study of choice behavior. Nature Neuroscience *8,* 1255-1262.**

**85. Coutureau,E., Dix,S.L., and Killcross,A.S. (2000). Involvement of the medial prefrontal cortex-basolateral amygdala pathway in fear related behaviour in rats. European Journal of Neuroscience *12,* 156.**

**86. Craig,A.D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. Nat.Rev.Neurosci. *3,* 655-666.**

**87. Craig,A.D. (2003). A new view of pain as a homeostatic emotion. Trends Neurosci. *26,* 303-307.**

**88. Critchley,H.D. and Rolls,E.T. (1996). Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. J.Neurophysiol. *75,* 1673-1686.**

**89. Daw,N.D. and Doya,K. (2006). The computational neurobiology of learning and reward. Curr.Opin.Neurobiol. *16,* 199-204.**

**90. Daw,N.D., Kakade,S., and Dayan,P. (2002). Opponent interactions between serotonin and dopamine. Neural Netw. *15,* 603-616.**

**91. Daw,N.D., Niv,Y., and Dayan,P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat.Neurosci. *8,* 1704-1711.**

**92. Daw,N.D. and Touretzky,D.S. (2002). Long-term reward prediction in TD models of the dopamine system. Neural Comput. *14,* 2567-2583.**

**93. Dayan,P. and Seymour,B. (2008). Value and actions in aversion. In Neuroeconomics: Decision making and the brain. Edited by Glimcher PW, Camerer CF, Fehr E, Poldrack RA. Elsevier.**

**94. Dayan,P. and Abbott LF. (2001). Theoretical Neuroscience: Compuational and Mathematical Modeling of Neural Systems. MIT Press.**

**95. Dayan,P. and Balleine,B.W. (2002). Reward, motivation, and reinforcement learning. Neuron *36,* 285-298.**

**96. Dayan,P. and Huys,Q.J. (2008). Serotonin, inhibition, and negative mood. PLoS.Comput.Biol. *4,* e4.**

**97. Dayan,P., Niv,Y., Seymour,B., and Daw,D. (2006). The misbehavior of value and the discipline of the will. Neural Netw.**

**98. de Araujo,I.E., Rolls,E.T., Velazco,M.I., Margot,C., and Cayeux,I. (2005). Cognitive modulation of olfactory processing. Neuron *46,* 671-679.**

99. De Martino,B., Kumaran,D., Seymour,B., and Dolan,R.J. (2006). Frames, biases, and rational decision-making in the human brain. Science *313,* 684-687.

100. de Quervain,D.J., Fischbacher,U., Treyer,V., Schellhammer,M., Schnyder,U., Buck,A., and Fehr,E. (2004a). The neural basis of altruistic punishment. Science *305,* 1254-1258.

101. de Quervain,D.J.F., Fischbacher,U., Treyer,V., Schelthammer,M., Schnyder,U., Buck,A., and Fehr,E. (2004b). The neural basis of altruistic punishment. Science *305,* 1254-1258.

102. De Villiers,P.A. (1974). The law of effect and avoidance: a quantitative relationship between response rate and shock-frequency reduction. J Exp Anal Behav 21, 223-235.

103. De Waal,F.B.M. (1998). Chimpanzee politics: Power and sex among apes. (Baltimore, MD: Johns Hopkins University Press).

104. Delgado,M.R., Nystrom,L.E., Fissell,C., Noll,D.C., and Fiez,J.A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. J.Neurophysiol. *84,* 3072-3077.

105. Delgado,M.R., Olsson,A., and Phelps,E.A. (2006). Extending animal models of fear conditioning to humans. Biol.Psychol. *73,* 39-48.

106. Denrell,J. (2007). Adaptive learning and risk taking. Psychological Review *114,* 177-187.

107. Denrell,J. and March,J.G. (2001). Adaptation as information restriction: The hot stove effect. Organization Science *12,* 523-538.

108. Dickinson,A. (1980). *Contemporary animal learning theory*. Cambridge, England: Cambridge University Press.

109. Dickinson,A. and Balleine,B.W. (2002). The role of learning in motivation.In: Gallistel, C.R., Editor, , 2002.Learning, Motivation and Emotion, Volume 3 of Steven's Handbook of Experimental Psychology (Third Edition ed.), John Wiley & Sons, New York in press.

110. Dickinson,A. and Dearing MF. (1979). Appetitive-aversive interactions and inhibitory processes. *In Mechanisms of Learning and Motivation.* Dickinson A and Boakes RA eds. Erlbaum, Hillsdale, NJ. 203-231.

111. Dinsmoor,J.A. (2001). Stimuli inevitably generated by behavior that avoids electric shock are inherently reinforcing. J.Exp.Anal.Behav. *75,* 311-333.

112. Doya,K. (2002). Metalearning and neuromodulation. Neural Netw. *15,* 495-506.

113. **Duvarci,S., Mamou,C.B., and Nader,K. (2006). Extinction is not a sufficient condition to prevent fear memories from undergoing reconsolidation in the basolateral amygdala. Eur.J.Neurosci. *24,* 249-260.**

114. **Elliott,R., Agnew,Z., and Deakin,J.F.W. (2008). Medial orbitofrontal cortex codes relative rather than absolute value of financial rewards in humans. European Journal of Neuroscience *27,* 2213-2218.**

115. **Elliott,R., Newman,J.L., Longe,O.A., and Deakin,J.F. (2003). Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: a parametric functional magnetic resonance imaging study. J.Neurosci. *23,* 303-307.**

116. **Estes,W.K. and Skinner,B.F. (1941). Some quantitative properties of anxiety. Journal of Experimental Psychology 29, 390-400.**

117. **Everitt,B.J., Parkinson,J.A., Olmstead,M.C., Arroyo,M., Robledo,P., and Robbins,T.W. (1999). Associative processes in addiction and reward. The role of amygdala-ventral striatal subsystems. Ann.N.Y.Acad.Sci. *877,* 412-438.**

118. **Fehr,E. and Fischbacher,U. (2003). The nature of human altruism. Nature *425,* 785-791.**

119. **Fehr,E. and Fischbacher,U. (2004a). Social norms and human cooperation. Trends Cogn Sci. *8,* 185-190.**

120. **Fehr,E. and Fischbacher,U. (2004b). Third-party punishment and social norms. Evolution and Human Behavior *25,* 63-87.**

121. **Fehr,E. and Gachter,S. (2000). Cooperation and punishment in public goods experiments. American Economic Review *90,* 980-994.**

122. **Fehr,E. and Gachter,S. (2002). Altruistic punishment in humans. Nature *415,* 137-140.**

123. **Fehr,E., Kirchsteiger,A., and Riedl,A. (1993). Does fairness prevent market clearing? An experimental investigation. Quarterly Journal of Economics *108,* 437-459.**

124. **Fendt,M. and Fanselow,M.S. (1999). The neuroanatomical and neurochemical basis of conditioned fear. Neurosci.Biobehav.Rev. *23,* 743-760.**

125. **Ferrari,E.A., Todorov,J.C., and Graeff,F.G. (1973). Nondiscriminated avoidance of shock by pigeons pecking a key. J.Exp.Anal.Behav 19, 211-218.**

126. **Fields HL and Price DD. in 'A Textbook of Pain'. (2005). Churchill livingstone, Edinburgh UK.  Ch 24.**

127. Fields,H. (2004). State-dependent opioid control of pain. Nat.Rev.Neurosci. *5,* 565-575.

128. Fields,H.L. (2000). Pain modulation: expectation, opioid analgesia and virtual pain. Prog.Brain Res. *122,* 245-253.

129. Fiorillo,C.D., Tobler,P.N., and Schultz,W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. Science *299,* 1898-1902.

130. Friston,K. (2003). Learning and inference in the brain. Neural Networks *16,* 1325-1352.

131. Friston,K.J., Tononi,G., Reeke,G.N., Jr., Sporns,O., and Edelman,G.M. (1994). Value-dependent selection in the brain: simulation in a synthetic neural model. Neuroscience *59,* 229-243.

132. Gadd,C.A., Murtra,P., De Felipe,C., and Hunt,S.P. (2003). Neurokinin-1 receptor-expressing neurons in the amygdala modulate morphine reward and anxiety behaviors in the mouse. J.Neurosci. *23,* 8271-8280.

133. Gallagher,H.L. and Frith,C.D. (2003). Functional imaging of 'theory of mind'. Trends Cogn Sci. *7,* 77-83.

134. Gallagher,H.L., Jack,A.I., Roepstorff,A., and Frith,C.D. (2002). Imaging the intentional stance in a competitive game. Neuroimage. *16,* 814-821.

135. Gallagher,M. and Chiba,A.A. (1996). The amygdala and emotion. Current Opinion in Neurobiology *6,* 221-227.

136. Gallagher,M. and Holland,P.C. (1994). The amygdala complex: multiple roles in associative learning and attention. Proc.Natl.Acad.Sci.U.S.A *91,* 11771-11776.

137. Gintis,H. (2000). Strong reciprocity and human sociality. J.Theor.Biol. *206,* 169-179.

138. Glascher,J. and Buchel,C. (2005a). Formal learning theory dissociates brain regions with different temporal integration. Neuron *47,* 295-306.

139. Glascher,J. and Buchel,C. 2005b. Formal learning theory dissociates brain regions with different temporal integration. Neuron *in press*.

140. Gottfried,J.A., O'Doherty,J., and Dolan,R.J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science *301,* 1104-1107.

141. Graeff,F.G. (2004). Serotonin, the periaqueductal gray and panic. Neuroscience and Biobehavioral Reviews *28,* 239-259.

142. Graeff,F.G., Guimaraes,F.S., DeAndrade,T.G.C.S., and Deakin,J.F.W. (1996). Role of 5-HT in stress, anxiety, and depression. Pharmacology Biochemistry and Behavior *54,* 129-141.

143. Gray,J.A. 1991. *The psychology of fear and stress*, volume 5 of *Problems in the behavioural sciences*. Cambridge University Press, Cambridge, UK, 2 edition.

144. Grossberg,S. (1984). Some normal and abnormal behavioral syndromes due to transmitter gating of opponent processes. Biol.Psychiatry *19,* 1075-1118.

145. Grossberg,S. (2000). The imbalanced brain: from normal behavior to schizophrenia. Biol.Psychiatry *48,* 81-98.

146. Gurerk,O., Irlenbusch,B., and Rockenbach,B. (2006). The competitive advantage of sanctioning institutions. Science *312,* 108-111.

147. H.Klüver and P.C.Bucy. 1939. Preliminary analysis of functions of the temporal lobes in monkeys. Arch Neurol Psychiatry 42, 979-997.

148. Hall,J., Parkinson,J.A., Connor,T.M., Dickinson,A., and Everitt,B.J. (2001). Involvement of the central nucleus of the amygdala and nucleus accumbens core in mediating Pavlovian influences on instrumental behaviour. Eur.J.Neurosci. *13,* 1984-1992.

149. Hampton,A.N., Adolphs,R., Tyszka,M.J., and O'Doherty,J.P. (2007). Contributions of the amygdala to reward expectancy and choice signals in human prefrontal cortex. Neuron *55,* 545-555.

150. Hampton,A.N., Bossaerts,P., and O'Doherty,J.P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J.Neurosci. *26,* 8360-8367.

151. Han,C.J., O'Tuathaigh,C.M., van Trigt,L., Quinn,J.J., Fanselow,M.S., Mongeau,R., Koch,C., and Anderson,D.J. (2003). Trace but not delay fear conditioning requires attention and the anterior cingulate cortex. Proc.Natl.Acad.Sci.U.S.A *100,* 13087-13092.

152. Harbaugh,W.T. (1998). The prestige motive for making charitable transfers. American Economic Review *88,* 277-282.

153. Haruno,M., Kuroda,T., Doya,K., Toyama,K., Kimura,M., Samejima,K., Imamizu,H., and Kawato,M. (2004). A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task. J.Neurosci. *24,* 1660-1665.

154. Hatfield,T., Han,J.S., Conley,M., Gallagher,M., and Holland,P. (1996). Neurotoxic lesions of basolateral, but not central, amygdala

**interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. J.Neurosci. *16,* 5256-5265.**

**155. Heidbreder,C.A., Hedou,G., and Feldon,J. (1999). Behavioral neurochemistry reveals a new functional dichotomy in the shell subregion of the nucleus accumbens. Progress in Neuro-Psychopharmacology & Biological Psychiatry *23,* 99-132.**

**156. Henderson,R.W. and Graham,J. (1979). Avoidance of Heat by Rats - Effects of Thermal Context on Rapidity of Extinction. Learning and Motivation *10,* 351-363.**

**157. Henrich,J. (2006). Cooperation, punishment, and the evolution of human institutions. Science *312,* 60-61.**

**158. Henrich,J., Boyd,R., Bowles,S., Camerer,C., Fehr,E., Gintis,H., and McElreath,R. (2001a). In search of Homo economicus: Behavioral experiments in 15 small-scale societies. American Economic Review *91,* 73-78.**

**159. Henrich,J., Boyd,R., Bowles,S., Camerer,C., Fehr,E., Gintis,H., and McElreath,R. (2001b). In search of Homo economicus: Behavioral experiments in 15 small-scale societies. American Economic Review *91,* 73-78.**

**160. Henrich,J., McElreath,R., Barr,A., Ensminger,J., Barrett,C., Bolyanatz,A., Cardenas,J.C., Gurven,M., Gwako,E., Henrich,N., Lesorogol,C., Marlowe,F., Tracer,D., and Ziker,J. (2006). Costly punishment across human societies. Science *312,* 1767-1770.**

**161. Hineline,P.N. 1977. Negative reinforcement and avoidance. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 364-414). Englewood Cliffs, NJ: Prentice Hall.**

**162. Hitchcott,P.K. and Phillips,G.D. (1998). Effects of intra amygdala R(+) 7-OH-DPAT on intraaccumbens d-amphetamine-associated learning - I. Pavlovian conditioning and II. Instrumental conditioning. Psychopharmacology *140,* 300-318.**

**163. Holland,P.C. and Gallagher,M. (1993). Amygdala central nucleus lesions disrupt increments, but not decrements, in conditioned stimulus processing. Behav.Neurosci. *107,* 246-253.**

**164. Holland,P.C. and Gallagher,M. (2003). Double dissociation of the effects of lesions of basolateral and central amygdala on conditioned stimulus-potentiated feeding and Pavlovian-instrumental transfer. Eur.J.Neurosci. *17,* 1680-1694.**

**165. Holland,P.C. and Gallagher,M. (2004). Amygdala-frontal interactions and reward expectancy. Curr.Opin.Neurobiol. *14,* 148-155.**

166. Holland,P.C., Han,J.S., and Winfield,H.M. (2002a). Operant and Pavlovian control of visual stimulus orienting and food-related behaviors in rats with lesions of the amygdala central nucleus. Behav.Neurosci. *116,* 577-587.

167. Holland,P.C., Hatfield,T., and Gallagher,M. (2001). Rats with basolateral amygdala lesions show normal increases in conditioned stimulus processing but reduced conditioned potentiation of eating. Behav.Neurosci. *115,* 945-950.

168. Holland,P.C., Petrovich,G.D., and Gallagher,M. (2002b). The effects of amygdala lesions on conditioned stimulus-potentiated eating in rats. Physiol Behav. *76,* 117-129.

169. Horvitz,J.C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. Neuroscience *96,* 651-656.

170. Hosokawa,T., Kato,K., Inoue,M., and Mikami,A. (2007). Neurons in the macaque orbitofrontal cortex code relative preference of both rewarding and aversive outcomes. Neuroscience Research *57,* 434-445.

171. Hsu,M., Bhatt,M., Adolphs,R., Tranel,D., and Camerer,C.F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. Science *310,* 1680-1683.

172. Hunt,S.P. and Mantyh,P.W. (2001). The molecular dynamics of pain control. Nat.Rev.Neurosci. *2,* 83-91.

173. Hutchinson,R.R., Azrin,N.H., and Hunt,G.M. (1968). Attack produced by intermittent reinforcement of a concurrent operant response. J.Exp.Anal.Behav. *11,* 489-495.

174. Huys,Q. and Dayan,P. 2008. A Bayesian formulation of behavioral control. submitted .

175. Ikemoto,S. and Panksepp,J. (1999). The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. Brain Res.Brain Res.Rev. *31,* 6-41.

176. Jensen,J., McIntosh,A.R., Crawley,A.P., Mikulis,D.J., Remington,G., and Kapur,S. (2003). Direct activation of the ventral striatum in anticipation of aversive stimuli. Neuron *40,* 1251-1257.

177. Jensen,J., Smith,A.J., Willeit,M., Crawley,A.P., Mikulis,D.J., Vitcu,I., and Kapur,S. (2006). Separate brain regions code for salience vs. valence during reward prediction in humans. Hum.Brain Mapp.

178. Johansen,J.P. and Fields,H.L. (2004). Glutamatergic activation of anterior cingulate cortex produces an aversive teaching signal. Nat.Neurosci. *7,* 398-403.

179. **Jones,A.K.P., Friston,K., and Frackowiak,R.S.J. (1992). Localization of Responses to Pain in Human Cerebral-Cortex. Science *255,* 215.**

180. **Julius,D. and Basbaum,A.I. (2001). Molecular mechanisms of nociception. Nature *413,* 203-210.**

181. **Kahneman,D. and Frederick,S. (2002) Representativeness revisited: attribute substitution in intuitive judgment. In: T. Gilovich et al., Editors, Heuristics and Biases: the Psychology of Intuitive Judgment, Cambridge University Press, pp. 49–81.**

182. **Kamin,L.J. 1968. 'Attention-like' processes in classical conditioning. In *Miami symposium on the prediction of behavior: aversive stimulation* (ed M.R. Jones) pp9-33. University of Miami Press.**

183. **Kamin,L.J., Black,A.H., and Brimer,C.J. (1963). Conditioned Suppression As A Monitor of Fear of Cs in Course of Avoidance Training. Journal of Comparative and Physiological Psychology *56,* 497-&.**

184. **Kersten,D. and Yuille,A. (2003). Bayesian models of object perception. Current Opinion in Neurobiology *13,* 150-158.**

185. **Killcross,S., Robbins,T.W., and Everitt,B.J. (1997). Different types of fear-conditioned behaviour mediated by separate nuclei within amygdala. Nature *388,* 377-380.**

186. **Kim,H., Shimojo,S., and O'Doherty,J.P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. Plos Biology *4,* 1453-1461.**

187. **Knoch,D., Pascual-Leone,A., Meyer,K., Treyer,V., and Fehr,E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. Science *314,* 829-832.**

188. **Knutson,B., Rick,S., Wimmer,G.E., Prelec,D., and Loewenstein,G. (2007a). Neural predictors of purchases. Neuron *53,* 147-156.**

189. **Knutson,B., Rick,S., Wirnmer,G.E., Prelec,D., and Loewenstein,G. (2007b). Neural predictors of purchases. Neuron *53,* 147-156.**

190. **Knutson,B., Westdorp,A., Kaiser,E., and Hommer,D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. Neuroimage. *12,* 20-27.**

191. **Knuttinen,M.G., Power,J.M., Preston,A.R., and Disterhoft,J.F. (2001). Awareness in classical differential eyeblink conditioning in young and aging humans. Behav.Neurosci. *115,* 747-757.**

192. **Konorski,J. 1967. Integrative Activity of the Brain: An Interdisciplinary Approach (Chicago: University of Chicago Press).**

**193. Koob,G.F., Caine,S.B., Parsons,L., Markou,A., and Weiss,F. (1997). Opponent process model and psychostimulant addiction. Pharmacol.Biochem.Behav. *57,* 513-521.**

**194. Laming,D. (1984). The Relativity of Absolute Judgements. British Journal of Mathematical & Statistical Psychology *37,* 152-183.**

**195. LeDoux,J. (1998). Fear and the brain: where have we been, and where are we going? Biol.Psychiatry *44,* 1229-1238.**

**196. LeDoux,J.E. (2000a). Emotion circuits in the brain. Annu.Rev.Neurosci. *23,* 155-184.**

**197. LeDoux,J.E. 2000b.The amygdala and emotion: A view through fear. In: J.P. Aggleton, Editor, The Amygdala: A functional analysis (2nd edn), Oxford University Press (2000), pp. 289-310.**

**198. Lee,D. (2006). Neural basis of quasi-rational decision making. Curr.Opin.Neurobiol. *16,* 191-198.**

**199. Lee,D. (2008). Game theory and neural basis of social decision making. Nat.Neurosci. *11,* 404-409.**

**200. Lengyel,M. and Dayan,P. 2007. Hippocampal contributions to control: The third way. NIPS.**

**200. Levita,L., Dalley,J.W., Robbins,T.W. (2002). Disruption of Pavlovian Contextual COnditioning by Excitotoxic Lesions of the Nucleus Accumbens Core. Behav.Neurosci. 116(4).539-52.**

**201. Levy,M. and Levy,H. (2002). Prospect theory: Much ado about nothing? Management Science *48,* 1334-1349.**

**202. Loewenstein,G. (2006). The pleasures and pains of information. Science *312,* 704-706.**

**203. Logothetis,N.K., Pauls,J., Augath,M., Trinath,T., and Oeltermann,A. (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature *412,* 150-157.**

**204. Mackintosh,N.J. 1983. *Conditioning and associative learning.* New York: Oxford University Press.**

**205. Mahadevan,S. 1996Average reward reinforcement learning: Foundations,**

**algorithms and empirical results. Machine Learning 22, 1-38.**

**206. Malkova,L., Gaffan,D., and Murray,E.A. (1997). Excitotoxic lesions of the amygdala fail to produce impairment in visual learning for auditory secondary reinforcement but interfere with reinforcer devaluation effects in rhesus monkeys. Journal of Neuroscience *17,* 6011-6020.**

207. **Mangel,M. and Clark,C.W. 1988. Dynamic Modelling in Behavioral Ecology. Princeton, New Jersey: Princeton University Press.**

208. **March,J.G. (1996). Learning to be risk averse. Psychological Review *103,* 309-319.**

209. **Maren,S. (2005). Synaptic mechanisms of associative memory in the amygdala. Neuron *47,* 783-786.**

210. **Maren,S. and Quirk,G.J. (2004). Neuronal signalling of fear memory. Nat.Rev.Neurosci. *5,* 844-852.**

211. **Markowitz,H. 1952. The utility of wealth. Journal of Political Economy 60, 151-158.**

212. **Marr,D. (1969). A theory of cerebellar cortex. J.Physiol *202,* 437-470.**

213. **Marr,D. (1970). A theory for cerebral neocortex. Proc.R.Soc.Lond B Biol.Sci. *176,* 161-234.**

214. **Marr,D. (1971). Simple memory: a theory for archicortex. Philos.Trans.R.Soc.Lond B Biol.Sci. *262,* 23-81.**

215. **Matsumoto,M. and Hikosaka,O. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. Nature *447,* 1111-1115.**

216. **McClure,S.M., Berns,G.S., and Montague,P.R. (2003). Temporal prediction errors in a passive learning task activate human striatum. Neuron *38,* 339-346.**

217. **McClure,S.M., Laibson,D.I., Loewenstein,G., and Cohen,J.D. (2004). Separate neural systems value immediate and delayed monetary rewards. Science *306,* 503-507.**

218. **McDannald,M.A., Saddoris,M.P., Gallagher,M., and Holland,P.C. (2005). Lesions of orbitofrontal cortex impair rats' differential outcome expectancy learning but not conditioned stimulus-potentiated feeding. J.Neurosci. *25,* 4626-4632.**

219. **Mesulam,M.M. and Mufson,E.J. (1982). Insula of the old world monkey. I. Architectonics in the insulo-orbito-temporal component of the paralimbic brain. J.Comp Neurol. *212,* 1-22.**

220. **Milad,M.R. and Quirk,G.J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. Nature *420,* 70-74.**

221. **Mineka,S. and Gino,A. (1980). Dissociation Between Conditioned Emotional Response and Extended Avoidance Performance. Learning and Motivation *11,* 476-502.**

222. **Mirenowicz,J. and Schultz,W. (1996). Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. Nature *379,* 449-451.**

223. **Mobbs,D., Marchant,J.L., Hassabis,D., Seymour,B., Tan,G., Gray,M., Petrovic,P., Dolan,R.J., and Frith,C.D. (2009). From threat to fear: the neural organization of defensive fear systems in humans. J.Neurosci. *29,* 12236-12243.**

224. **Mobbs,D., Petrovic,P., Marchant,J.L., Hassabis,D., Weiskopf,N., Seymour,B., Dolan,R.J., and Frith,C.D. (2007). When fear is near: Threat imminence elicits prefrontal-periaqueductal gray shifts in humans. Science *317,* 1079-1083.**

225. **Moll,J., Zahn,R., Oliveira-Souza,R., Krueger,F., and Grafman,J. (2005). Opinion: the neural basis of human moral cognition. Nat.Rev.Neurosci. *6,* 799-809.**

226. **Montague,P.R. and Berns,G.S. (2002). Neural economics and the biological substrates of valuation. Neuron *36,* 265-284.**

227. **Montague,P.R., Dayan,P., and Sejnowski,T.J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J.Neurosci. *16,* 1936-1947.**

228. **Montague,P.R., Hyman,S.E., and Cohen,J.D. (2004). Computational roles for dopamine in behavioural control. Nature *431,* 760-767.**

229. **Morris,G., Nevet,A., Arkadir,D., Vaadia,E., and Bergman,H. (2006). Midbrain dopamine neurons encode decisions for future action. Nat.Neurosci. *9,* 1057-1063.**

230. **Morris,J.S., Ohman,A., and Dolan,R.J. (1998). Conscious and unconscious emotional learning in the human amygdala. Nature *393,* 467-470.**

231. **Mowrer,O.H. 1947.On the dual nature of learning: A re-interpretation of" conditioning" and" problem-solving. Harvard Educational Review 17, 102-148.**

232. **Mufson,E.J., Mesulam,M.M., and Pandya,D.N. (1981). Insular interconnections with the amygdala in the rhesus monkey. Neuroscience *6,* 1231-1248.**

233. **Murray,E.A. (2007). The amygdala, reward and emotion. Trends Cogn Sci. *11,* 489-497.**

234. **Nakahara,H., Itoh,H., Kawagoe,R., Takikawa,Y., and Hikosaka,O. (2004). Dopamine neurons can represent context-dependent prediction error. Neuron *41,* 269-280.**

**235. Nieuwenhuis,S., Heslenfeld,D.J., von Geusau,N.J., Mars,R.B., Holroyd,C.B., and Yeung,N. (2005). Activity in human reward-sensitive brain areas is strongly context dependent. Neuroimage. *25,* 1302-1309.**

**236. Nitschke,J.B., Sarinopoulos,I., Mackiewicz,K.L., Schaefer,H.S., and Davidson,R.J. (2006). Functional neuroanatomy of aversion and its anticipation. Neuroimage. *29,* 106-116.**

**237. Niv,Y., Daw,N.D., Joel,D., and Dayan,P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology *191,* 507-520.**

**238. Niv,Y., Duff,M.O., and Dayan,P. (2005). Dopamine, uncertainty and TD learning. Behav.Brain Funct. *1,* 6.**

**239. Niv,Y., Joel,D., Meilijson,I., and Ruppin,E. (2002). Evolution of reinforcement learning in foraging bees: a simple explanation for risk averse behavior. Neurocomputing *44,* 951-956.**

**240. O'Doherty,J., Kringelbach,M.L., Rolls,E.T., Hornak,J., and Andrews,C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. Nat.Neurosci. *4,* 95-102.**

**241. O'Doherty,J.P., Dayan,P., Friston,K., Critchley,H., and Dolan,R.J. (2003). Temporal difference models and reward-related learning in the human brain. Neuron *38,* 329-337.**

**242. Ohman,A. and Soares,J.J. (1998). Emotional conditioning to masked stimuli: expectancies for aversive outcomes following nonrecognized fear-relevant stimuli. J.Exp.Psychol.Gen. *127,* 69-82.**

**243. Ostlund,S.B. and Balleine,B.W. (2007). Orbitofrontal cortex mediates outcome encoding in pavlovian but not instrumental conditioning. Journal of Neuroscience *27,* 4819-4825.**

**244. Overmier,J.B., Bull,J.A., and Trapold,M.A. (1971b). Discriminative Cue Properties of Different Fears and Their Role in Response Selection in Dogs. Journal of Comparative and Physiological Psychology *76,* 478-&.**

**245. Overmier,J.B., Bull,J.A., and Trapold,M.A. (1971a). Discriminative Cue Properties of Different Fears and Their Role in Response Selection in Dogs. Journal of Comparative and Physiological Psychology *76,* 478-&.**

**246. Padoa-Schioppa,C. and Assad,J.A. (2006). Neurons in the orbitofrontal cortex encode economic value. Nature *441,* 223-226.**

**247. Padoa-Schioppa,C. and Assad,J.A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. Nature Neuroscience *11,* 95-102.**

248. **Padoa-Schioppa,C., Jandolo,L., and Visalberghi,E. (2006). Multistage mental process for economic choice in capuchins. Cognition *99,* B1-B13.**

249. **Pagnoni,G., Zink,C.F., Montague,P.R., and Berns,G.S. (2002). Activity in human ventral striatum locked to errors of reward prediction. Nat.Neurosci. *5,* 97-98.**

250. **Parkinson,J.A., Dalley,J.W., Cardinal,R.N., Bamford,A., Fehnert,B., Lachenal,G., Rudarakanchana,N., Halkerston,K.M., Robbins,T.W., and Everitt,B.J. (2002). Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function. Behavioural Brain Research *137,* 149-163.**

251. **Parkinson,J.A., Willoughby,P.J., Robbins,T.W., and Everitt,B.J. (2000). Disconnection of the anterior cingulate cortex and nucleus accumbens core impairs Pavlovian approach behavior: Further evidence for limbic cortical-ventral striatopallidal systems. Behavioral Neuroscience *114,* 42-63.**

252. **Paton,J.J., Belova,M.A., Morrison,S.E., and Salzman,C.D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature *439,* 865-870.**

253. **Paulus,M.P. and Stein,M.B. (2006). An insular view of anxiety. Biological Psychiatry *60,* 383-387.**

254. **Pessiglione,M., Seymour,B., Flandin,G., Dolan,R.J., and Frith,C.D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature *442,* 1042-1045.**

255. **Petrovic,P., Kalso,E., Petersson,K.M., and Ingvar,M. (2002). Placebo and opioid analgesia-- imaging a shared neuronal network. Science *295,* 1737-1740.**

256. **Petrovich,G.D., Holland,P.C., and Gallagher,M. (2005). Amygdalar and prefrontal pathways to the lateral hypothalamus are activated by a learned cue that stimulates eating. J.Neurosci. *25,* 8295-8302.**

257. **Petrovich,G.D., Setlow,B., Holland,P.C., and Gallagher,M. (2002). Amygdalo-hypothalamic circuit allows learned cues to override satiety and promote eating. J.Neurosci. *22,* 8748-8753.**

258. **Phelps,E.A. and LeDoux,J.E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. Neuron *48,* 175-187.**

259. **Plassmann,H., O'Doherty,J., and Rangel,A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. Journal of Neuroscience *27,* 9984-9988.**

**260. Plassmann,H., O'Doherty,J., Shiv,B., and Rangel,A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. Proceedings of the National Academy of Sciences of the United States of America *105,* 1050-1054.**

**261. Ploghaus,A., Becerra,L., Borras,C., and Borsook,D. (2003). Neural circuitry underlying pain modulation: expectation, hypnosis, placebo. Trends Cogn Sci. *7,* 197-200.**

**262. Ploghaus,A., Tracey,I., Clare,S., Gati,J.S., Rawlins,J.N., and Matthews,P.M. (2000). Learning about pain: the neural substrate of the prediction error for aversive events. Proc.Natl.Acad.Sci.U.S.A *97,* 9281-9286.**

**263. Ploghaus,A., Tracey,I., Gati,J.S., Clare,S., Menon,R.S., Matthews,P.M., and Rawlins,J.N. (1999). Dissociating pain from its anticipation in the human brain. Science *284,* 1979-1981.**

**264. Poldrack,R.A. (2006). Can cognitive processes be inferred from neuroimaging data? Trends Cogn Sci. *10,* 59-63.**

**265. Price,D.D. 1999. Psychological Mechanisms of Pain and Analgesia. IASP press, Seattle USA.**

**266. Puterman,M. 1994. Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons, Inc. New York, NY, USA.**

**267. Quirk,G.J., Repa,C., and LeDoux,J.E. (1995). Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: parallel recordings in the freely behaving rat. Neuron *15,* 1029-1039.**

**268. Raby,C.R., Alexis,D.M., Dickinson,A., and Clayton,N.S. (2007). Planning for the future by western scrub-jays. Nature *445,* 919-921.**

**269. Reeve,H.K. (1992). Queen Activation of Lazy Workers in Colonies of the Eusocial Naked Mole-Rat. Nature *358,* 147-149.**

**270. Rescorla RA. 1971. Variation in the effectiveness of reinforcement and non-reinforcement following proir inhibitory conditioning. Learn.Motiv. 2, 113-123.**

**271. Rescorla,R.A. and & Wagner,A.R. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II* (pp. 64-99). New York: Appleton-Century-Crofts.**

**272. Rescorla,R.A. (1969). Pavlovian Conditioned Inhibition. Psychological Bulletin *72,* 77-&.**

273. **Reynolds,S.M. and Berridge,K.C. (2001). Fear and feeding in the nucleus accumbens shell: rostrocaudal segregation of GABA-elicited defensive behavior versus eating behavior. J.Neurosci. *21,* 3261-3270.**

274. **Reynolds,S.M. and Berridge,K.C. (2002). Positive and negative motivation in nucleus accumbens shell: bivalent rostrocaudal gradients for GABA-elicited eating, taste "liking"/"disliking" reactions, place preference/avoidance, and fear. J.Neurosci. *22,* 7308-7320.**

275. **Reynolds,S.M. and Berridge,K.C. (2003). Glutamate motivational ensembles in nucleus accumbens: rostrocaudal shell gradients of fear and feeding. Eur.J.Neurosci. *17,* 2187-2200.**

276. **Riess,D. 1971. Shuttleboxes, Skinner boxes, and Sidman aoidance in rats: acquistion and terminal performance as a function of response topography. Psychonom.Sci. 25, 283-286.**

277. **Robbins,T.W. and Crockett,M.J. 2009. The role of serotonin in impulsivity and compulsivity: Comparative studies in experimental animals and humans. In: The behavioral neurobiology of serotonin, Eds: Muller, C.P. & Jacobs, B.**

278. **Robinson,T.N., Borzekowski,D.L.G., Matheson,D.M., and Kraemer,H.C. (2007). Effects of fast food branding on young children's taste preferences. Archives of Pediatrics & Adolescent Medicine *161,* 792-797.**

279. **Rockenbach,B. and Milinski,M. (2006). The efficient interaction of indirect reciprocity and costly punishment. Nature *444,* 718-723.**

280. **Roe,R.M., Busemeyer,J.R., and Townsend,J.T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. Psychological Review *108,* 370-392.**

281. **Roesch,M.R. and Olson,C.R. (2004). Neuronal activity related to reward value and motivation in primate frontal cortex. Science *304,* 307-310.**

282. **Rogan,M.T., Leon,K.S., Perez,D.L., and Kandel,E.R. (2005). Distinct neural signatures for safety and danger in the amygdala and striatum of the mouse. Neuron *46,* 309-320.**

283. **Roitman,M.F., Wheeler,R.A., and Carelli,R.M. (2005). Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. Neuron *45,* 587-597.**

284. **Rolls,E.T. (2000). The orbitofrontal cortex and reward. Cereb.Cortex *10,* 284-294.**

**285. Romo,R. and Schultz,W. (1989). Somatosensory input to dopamine neurones of the monkey midbrain: responses to pain pinch under anaesthesia and to active touch in behavioural context. Prog.Brain Res. *80,* 473-478.**

**286. Russchen,F.T., Bakst,I., Amaral,D.G., and Price,J.L. (1985). The amygdalostriatal projections in the monkey. An anterograde tracing study. Brain Res. *329,* 241-257.**

**287. Saddoris,M.P., Gallagher,M., and Schoenbaum,G. (2005). Rapid associative encoding in basolateral amygdala depends on connections with orbitofrontal cortex. Neuron *46,* 321-331.**

**288. Sanfey,A.G., Rilling,J.K., Aronson,J.A., Nystrom,L.E., and Cohen,J.D. (2003). The neural basis of economic decision-making in the Ultimatum Game. Science *300,* 1755-1758.**

**289. Sarinopoulos,I., Dixon,G.E., Short,S.J., Davidson,R.J., and Nitschke,J.B. (2006). Brain mechanisms of expectation associated with insula and amygdala response to aversive taste: implications for placebo. Brain Behav.Immun. *20,* 120-132.**

**290. Satoh,T., Nakai,S., Sato,T., and Kimura,M. (2003). Correlated coding of motivation and outcome of decision by dopamine neurons. J.Neurosci. *23,* 9913-9923.**

**291. Schoenbaum,G. and Setlow,B. (2003). Lesions of Nucleus Accumbens Disrupt Learning about Aversive Outcomes. J.Neurosci. *23,* 9833-9841.**

**292. Schoenbaum,G., Setlow,B., Saddoris,M.P., and Gallagher,M. (2003). Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. Neuron *39,* 855-867.**

**293. Schull,J. 1979. A conditioned opponent theory of Pavlovian conditioning and habituation. In: G. Bower (Ed.), The Psychology of Learning and Motivation, New York:Academic Press. 57-90.**

**294. Schultz,W. (2000). Multiple reward signals in the brain. Nat.Rev.Neurosci. *1,* 199-207.**

**295. Schultz,W., Dayan,P., and Montague,P.R. (1997). A neural substrate of prediction and reward. Science *275,* 1593-1599.**

**296. Schwartz,A. 1993. A reinforcement learning method for maximizing undiscounted rewards. In Proceedings of the Tenth International Conference on Machine Learning (pp. 298-305). San Mateo, CA: Morgan Kaufmann.**

**297. Schwartz,O., Sejnowski,T.J., and Dayan,P. (2009). Perceptual organization in the tilt illusion. Journal of Vision *9*.**

298. **Selden,N.R., Everitt,B.J., Jarrard,L.E., and Robbins,T.W. (1991). Complementary roles for the amygdala and hippocampus in aversive conditioning to explicit and contextual cues. Neuroscience *42,* 335-350.**

299. **Setlow,B., Schoenbaum,G., and Gallagher,M. (2003). Neural encoding in ventral striatum during olfactory discrimination learning. Neuron *38,* 625-636.**

300. **Seymour,B. (2006). Carry on eating: Neural pathways mediating conditioned Potentiation of feeding. Journal of Neuroscience *26,* 1061-1062.**

301. **Seymour,B., Daw,N., Dayan,P., Singer,T., and Dolan,R. (2007a). Differential encoding of losses and gains in the human striatum. J.Neurosci. *27,* 4826-4831.**

302. **Seymour,B., O'Doherty,J.P., Dayan,P., Koltzenburg,M., Jones,A.K., Dolan,R.J., Friston,K.J., and Frackowiak,R.S. (2004). Temporal difference models describe higher-order learning in humans. Nature *429,* 664-667.**

303. **Seymour,B., O'Doherty,J.P., Koltzenburg,M., Wiech,K., Frackowiak,R., Friston,K., and Dolan,R. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. Nat.Neurosci. *8,* 1234-1240.**

304. **Seymour,B., Singer,T., and Dolan,R. (2007b). The neurobiology of punishment. Nature Reviews Neuroscience *8,* 300-311.**

305. **Seymour,B., Yoshida,W., and Dolan,R. (2009). Altruistic learning. Front Behav.Neurosci. *3,* 23.**

306. **Shinada,M., Yamagishi,T., and Ohmura,Y. (2004). False friends are worse than bitter enemies: "Altruistic" punishment of in-group members. Evolution and Human Behavior *25,* 379-393.**

307. **Shiv,B., Carmon,Z., and Ariely,D. (2005). Placebo effects of marketing actions: Consumers may get what they pay for. Journal of Marketing Research *42,* 383-393.**

308. **Singer,T., Seymour,B., O'Doherty,J.P., Stephan,K.E., Dolan,R.J., and Frith,C.D. (2006). Empathic neural responses are modulated by the perceived fairness of others. Nature *439,* 466-469.**

309. **Sloman,S.A. (1996). The empirical case for two systems of reasoning. Psychological Bulletin *119,* 3-22.**

310. **Small,D.M., Zatorre,R.J., Dagher,A., Evans,A.C., and Jones-Gotman,M. (2001). Changes in brain activity related to eating chocolate: from pleasure to aversion. Brain *124,* 1720-1733.**

311. **Smith,A.J., Becker,S., and Kapur,S. (2005). A computational model of the functional role of the ventral-striatal D2 receptor in the expression of previously acquired behaviors. Neural Comput. *17,* 361-395.**

312. **Smith,A. [1759] *The Theory of Moral Sentiments*. Edited by D.D. Raphael and A.L. Macfie. Oxford: Oxford University Press. 1976.**

313. **Solomon,R.L. (1980a). Recent experiments testing an opponent-process theory of acquired motivation. Acta Neurobiol.Exp.(Wars.) *40,* 271-289.**

314. **Solomon,R.L. (1980b). The opponent-process theory of acquired motivation: the costs of pleasure and the benefits of pain. Am.Psychol. *35,* 691-712.**

315. **Solomon,R.L. and Corbit,J.D. (1974). An opponent-process theory of motivation. I. Temporal dynamics of affect. Psychol.Rev. *81,* 119-145.**

316. **Solomon,R.L., Turner,L.H., and Lessac,M.S. (1968). Some effects of delay of punishment on resistance to temptation in dogs. J.Pers.Soc.Psychol. *8,* 233-238.**

317. **Stalnaker,T.A., Franz,T.M., Singh,T., and Schoenbaum,G. (2007). Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. Neuron *54,* 51-58.**

318. **Starr,M.D. and Mineka,S. (1977). Determinants of Fear Over Course of Avoidance-Learning. Learning and Motivation *8,* 332-350.**

319. **Stefanovic,B., Warnking,J.M., and Pike,G.B. (2004). Hemodynamic and metabolic responses to neuronal inhibition. Neuroimage *22,* 771-778.**

320. **Stevens,J.R. (2004). The selfish nature of generosity: harassment and food sharing in primates. Proc.Biol.Sci. *271,* 451-456.**

321. **Stevens,J.R. and Hauser,M.D. (2004). Why be nice? Psychological constraints on the evolution of cooperation. Trends in Cognitive Sciences *8,* 60-65.**

322. **Stewart,N., Chater,N., and Brown,G.D.A. (2006). Decision by sampling. Cognitive Psychology *53,* 1-26.**

323. **Sugrue,L.P., Corrado,G.S., and Newsome,W.T. (2005). Choosing the greater of two goods: neural currencies for valuation and decision making. Nat.Rev.Neurosci. *6,* 363-375.**

324. **Suri,R.E. and Schultz,W. (2001). Temporal difference model reproduces anticipatory neural activity. Neural Comput. *13,* 841-862.**

**325. Sutton RS and Barto AG. 1990. Time-derivative models of Pavlovian reinforcement. *In Learning and Computational Neuroscience: Foundations of Adaptive Networks.* Gabriel M and Moore J eds. MIT press, Cambridge, MA. 497-537.**

**326. Sutton,R.S. and Barto,A.G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. Psychol.Rev. *88,* 135-170.**

**327. Sutton,R.S. and Barto,A.G. 1998. *Reinforcement Learning. An introduction.* MIT press (Cambridge MA).**

**328. Tanaka,S.C., Doya,K., Okada,G., Ueda,K., Okamoto,Y., and Yamawaki,S. (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. Nat.Neurosci. *7,* 887-893.**

**329. Tanaka,S.C., Samejima,K., Okada,G., Ueda,K., Okamoto,Y., Yamawaki,S., and Doya,K. (2006). Brain mechanism of reward prediction under predictable and unpredictable environmental dynamics. Neural Netw. *19,* 1233-1241.**

**330. Tanimoto,H., Heisenberg,M., and Gerber,B. (2004). Experimental psychology: event timing turns punishment to reward. Nature *430,* 983.**

**331. Thomson,J.J. 1986. Rights, Restitution and Risk (Harvard Univ. Press, Cambridge), pp. 94-116.**

**332. Thorndike,E.L. 1911. Animal Intelligence. New York: Macmillan.**

**333. Tobler,P.N., Fiorillo,C.D., and Schultz,W. (2005b). Adaptive coding of reward value by dopamine neurons. Science *307,* 1642-1645.**

**334. Tobler,P.N., Fiorillo,C.D., and Schultz,W. (2005a). Adaptive coding of reward value by dopamine neurons. Science *307,* 1642-1645.**

**335. Todorov,E. (2004). Optimality principles in sensorimotor control. Nat.Neurosci. *7,* 907-915.**

**336. Tolman,E.C. 1932. Purposive Behavior in Animals and Men (Century, New York, 1932).**

**337. Tom,S.M., Fox,C.R., Trepel,C., and Poldrack,R.A. 2007. The Neural Basis of Loss Aversion in Decision-Making Under Risk. Science 315, 515-518.**

**338. Tomlin,D., Kayali,M.A., King-Casas,B., Anen,C., Camerer,C.F., Quartz,S.R., and Montague,P.R. (2006). Agent-specific responses in the cingulate cortex during economic exchanges. Science *312,* 1047-1050.**

339. **Tremblay,L. and Schultz,W. (1999). Relative reward preference in primate orbitofrontal cortex. Nature *398,* 704-708.**

340. **Trivers,R. (1971). The evolution of reciprocal altruism. Quarterly Review of Biology *46,* 35-57.**

341. **Tversky,A. and Kahneman,D. (1981a). The Framing of Decisions and the Psychology of Choice. Science *211,* 453-458.**

342. **Tversky,A. and Kahneman,D. (1981b). The framing of decisions and the psychology of choice. Science *211,* 453-458.**

343. **Ulrich,R.E. and Azrin,N.H. (1962). Reflexive fighting in response to aversive stimulation. J.Exp.Anal.Behav. *5,* 511-520.**

344. **Ungless,M.A. (2004). Dopamine: the salient issue. Trends Neurosci. *27,* 702-706.**

345. **Ungless,M.A., Magill,P.J., and Bolam,J.P. (2004). Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. Science *303,* 2040-2042.**

346. **Ursu,S. and Carter,C.S. (2005). Outcome representations, counterfactual comparisons and the human orbitofrontal cortex: implications for neuroimaging studies of decision-making. Brain Res.Cogn Brain Res. *23,* 51-60.**

347. **Vlaev,I. and Chater,N. (2006). Game relativity: how context influences strategic decision making. J.Exp.Psychol.Learn.Mem.Cogn *32,* 131-149.**

348. **Waber,R.L., Shiv,B., Carmon,Z., and Ariely,D. (2008). Commercial features of placebo and therapeutic efficacy. Jama-Journal of the American Medical Association *299,* 1016-1017.**

349. **Wager,T.D., Rilling,J.K., Smith,E.E., Sokolik,A., Casey,K.L., Davidson,R.J., Kosslyn,S.M., Rose,R.M., and Cohen,J.D. (2004). Placebo-induced changes in FMRI in the anticipation and experience of pain. Science *303,* 1162-1167.**

350. **Walker,S.C., Robbins,T.W., and Roberts,A.C. (2009). Response disengagement on a spatial self-ordered sequencing task: effects of regionally selective excitotoxic lesions and serotonin depletion within the prefrontal cortex. J.Neurosci. *29,* 6033-6041.**

351. **Walters,G.C. and Grusec,J.E. 1977. Punishment. ed. Freeman,W.H. San Francisco.**

352. **Watkins,C.J.C.H. and Dayan,P. (1992). Q-Learning. Machine Learning *8,* 279-292.**

353. **WEISKRANTZ,L. (1956). Behavioral changes associated with ablation of the amygdaloid complex in monkeys. J.Comp Physiol Psychol. *49,* 381-391.**

354. **Williams,D.R. and Williams,H. (1969). Auto-Maintenance in Pigeon - Sustained Pecking Despite Contingent Non-Reinforcement. Journal of the Experimental Analysis of Behavior *12,* 511-&.**

355. **Wilson,D.I. and Bowman,E.M. (2005). Rat nucleus accumbens neurons predominantly respond to the outcome-related properties of conditioned stimuli rather than their behavioral-switching properties. J.Neurophysiol. *94,* 49-61.**

356. **Wise,R.A. (2004). Dopamine, learning and motivation. Nat.Rev.Neurosci. *5,* 483-494.**

357. **Xiao,E. and Houser,D. (2005). Emotion expression in human punishment behavior. Proceedings of the National Academy of Sciences of the United States of America *102,* 7398-7401.**

358. **Yacubian,J., Glascher,J., Schroeder,K., Sommer,T., Braus,D.F., and Buchel,C. (2006). Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. J.Neurosci. *26,* 9530-9537.**

359. **Yamagishi,T. (1986). The Provision of A Sanctioning System As A Public Good. Journal of Personality and Social Psychology *51,* 110-116.**

360. **Yamagishi,T. and Sato,K. 1986. Motivational basis of the public goods problem. Journal of Personality and Social Psychology 50, 67-73.**

361. **Yoshida,W. and Ishii,S. (2006). Resolution of uncertainty in prefrontal cortex. Neuron *50,* 781-789.**

362. **Zink,C.F., Pagnoni,G., Martin,M.E., Dhamala,M., and Berns,G.S. (2003). Human striatal response to salient nonrewarding stimuli. J.Neurosci. *23,* 8092-8097.**